TEXTILE FINGERPRINTING FOR DISMOUNT ANALYSIS IN THE VISIBLE, NEAR, AND SHORTWAVE INFRARED DOMAIN

THESIS

Jennifer S. Yeom, Second Lieutenant, USAF

AFIT-ENG-14-M-86

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-14-M-86

TEXTILE FINGERPRINTING FOR DISMOUNT ANALYSIS IN THE VISIBLE, NEAR, AND SHORTWAVE INFRARED DOMAIN

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Electrical Engineering

Jennifer S. Yeom, B.S.E.E.

Second Lieutenant, USAF

March 2014

AFIT-ENG-14-M-86

TEXTILE FINGERPRINTING FOR DISMOUNT ANALYSIS IN THE VISIBLE, NEAR, AND SHORTWAVE INFRARED DOMAIN

Jennifer S. Yeom, B.S.E.E.
Second Lieutenant, USAF

Approved:

//signed//	6 Mar 2014
Lt Col Jeffrey D. Clark, PhD (Chairman)	Date
//signed//	10 Mar 2014
Maj Brian G. Woolley, PhD (Member)	Date
//signed//	6 Mar 2014
Richard K. Martin, PhD (Member)	Date

AFIT-ENG-14-M-86

Abstract

The ability to accurately and quickly locate an individual, or a dismount, is useful in a variety of situations and environments. A dismount's characteristics such as their gender, height, weight, build, and ethnicity could be used as discriminating factors. Hyperspectral imaging (HSI) is widely used in efforts to identify materials based on their spectral signatures. More specifically, HSI has been used for skin and clothing classification and detection. The ability to detect textiles (clothing) provides a discriminating factor that can aid in a more comprehensive detection of dismounts.

This thesis demonstrates the application of several feature selection methods (*i.e.,* support vector machines with recursive feature reduction, fast correlation based filter) in highly dimensional data collected from a spectroradiometer. The classification of the data is accomplished with the selected features and artificial neural networks. A model for uniquely identifying (fingerprinting) textiles are designed, where color and composition ard deternimed in order to fingerprint a specific textile. An artificial neural network is created based on the knowledge of the textile's color and composition, providing a uniquely identifying fingerprinting of a textile. Results show 100% accuracy for color and composition classification, and 98% accuracy for the overall textile fingerprinting process.

Acknowledgments

I would like to thank my advisor, LtCol Jeffrey Clark, for the amazing amount of support and guidance he has given me through out my time here at AFIT. Many thanks to my research group, especially Lindsay Cain and Alice Chen, for always having answers and solutions to all the questions and requests I've had along the way. And lastly, thanks to my family and friends, as I couldn't have done this without their support every step of the way.

Jennifer S. Yeom

Table of Contents

Page

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
AFIT	Air Force Institute of Technology
USAF	United States Air Force
VIS	Visible
NIR	Near Infrared
SWIR	Short-Wave Infrared
DOI	Dismount of Interest
ISR	Intelligence, Surveillance, Reconnaissance
HSI	Hyperspectral Imaging
FOI	Fabric of Interest
FCBF	Fast Correlation Based Filter
SVM	Support Vector Machine
RBF	Radial Basis Function
MLP	Multilayer Perceptron
ANN	Artificial Neural Network
SVM-RFE	Support Vector Machines with the aid of Recursive Feature Elimination
RGB	Red Green Blue
SAR	Synthetic Aperture Radar
HOG	Histograms of Oriented Gradients
NDSI	Normalized Difference Skin Index
linSVM	Linear Support Vector Machine
PCA	Principal Component Analysis
LVQ	Learning Vector Quantization
GRLVQ	Generalized Relevance Learning Vector Quantization

Abbreviation	Definition
PDF	Probability Distribution Function
CFAR	Constant False Alarm Rate
AD	Anomaly Detector
SU	Symmetrical Uncertainty
aLDA	augmented Latent Dirichlet Allocation
HMM	Hidden Markov Model
MSE	Mean Squared Error
EWA	Equal Weighted Accuracy
FN	False Negative
FP	False Positive

TEXTILE FINGERPRINTING FOR DISMOUNT ANALYSIS IN THE VISIBLE, NEAR, AND SHORTWAVE INFRARED DOMAIN

I. Introduction

DISMOUNT detection is the process of locating and identifying a person based on prior knowledge of the target. A dismount is defined as an individual who is on the ground, traveling on foot, and not in a vehicle [30]. Current technology has enhanced the capability to identify and track dismounts [30]. Applications of dismount detection include search and rescue, surveillance, and target identification [14].

Prior knowledge of a target is necessary for accurate detection and identification. When searching for a dismount of interest (DOI), it is crucial to know certain inherent characteristics such as height, weight, gender, age and ethnicity. Other pertinent information, *e.g.,* the clothing currently worn by a dismount, can provide valuable insight in detection and identification. Textile fingerprinting can be useful as a discriminating factor in identifying a DOI. This chapter introduces the problem statement of the thesis as well as the scope and limitations. The methodology, data, and results are also briefly discussed.

1.1 Background

Some of the most common human identification techniques include iris analysis, face, fingerprint and voice recognition [33, 63]. These biological features are popular due to their individual uniqueness; however, these identification techniques require either a close up image or a physical sample. The methods required for the previously mentioned identification techniques may not be possible, especially if the DOI is avoiding detection.

Recent studies have explored the detection of dismounts based on the discrimination of hair, skin and clothing with hyperspectral data [3]. Collecting information surreptitiously can be advantageous to the detectors (people trying to detect a dismount) in certain endeavors. Detecting a dismount passively prevents the dismount's awareness of detection and may be the only possible means to collect the information. An advantage passive detection identification methods have over others is that the distance between the target and the sensor can be nondeterministic. As long as the sensor has visual contact with the target, information can be gained and used for the purpose of detecting or identifying a dismount.

To illustrate the usefulness of dismount identification, two examples will be discussed. The first example involves a fugitive running from authorities. This fugitive is actively avoiding being detected. Visual identification or face recognition is unavailable because the fugitive has a hooded sweatshirt or a baseball cap on. Skin detection is also complicated by gloves and long sleeves as well as pants. Physical attributes, such as height and build, are negligable because the fugitive is in a highly crowded area. Assuming that the detectors know what the fugitive is wearing, the detector will be able to be identify the fugitive by classification of textiles. Because textile identification is achieved through passive methods, the fugitive will not be aware that he or she has been detected.

The second example involves a lost hiker in the woods. In this case, the dismount is not actively avoiding detection; however, it is highly likely that the terrain and the surroundings will complicate traditional detection methods. Clothing detection will enable the detectors to separate that individual from his or her surrounding environment, even if the dismount is immobile.

Accurate identification of dismounts is useful for multiple applications including intelligence, surveillance and reconnaissance (ISR). The motivation behind textile detection and identification is based on the fact that the largest imaged surface area of

dismounts is typically clothing. Two obvious characteristics of textiles are their color and material composition (textile make). A red 100% cotton shirt and a blue 100% cotton shirt are separable based on their color alone. A green 100% cotton shirt and a green 100% polyester shirt are differentiable due to their material composition. However, color and composition alone do not give enough information to uniquely identify (fingerprint) a subject.

Dismount detection using clothing is complicated when two dismounts are wearing the same clothing. The spectral signatures of two textiles of identical color and composition are non-separable. An example where multiple dismounts could exist wearing the same colored and type of clothing at the same time would be a sporting event. Fans tend to wear identical jerseys of the home or away teams, mass produced from only a few manufacturers. Two possible scenes for dismount detection are portrayed in Figure 1.1.

(a) (b)

Figure 1.1: Two example scenes of dismount detection. The image on the left (a) shows a crowded street view of the winter, where the only visible skin of the dismounts is the face [6]. The image on the right (b) shows dismounts at a sporting event, all wearing similar colors and jerseys manufactured from a few of the same companies [21].

The goal of this thesis is to identify various materials and blends of textiles as well as to uniquely identify textiles of the same type. Even if two articles of clothing are

manufactured by the same company and appear identical, the difference exists in the maintenance, maintainability, and environmental factors of the two articles. These differences will manifest in a distinguishable difference of their spectral signature. This thesis focuses on fingerprinting textiles for the purpose of the detection and identification of dismounts.

1.1.1 Hyperspectral Data.

Detection of skin and clothing has been accomplished through the use of hyperspectral imaging (HSI) [14, 55]. HSI collects high resolution of a material's spectral radiance over a large spectral range [7]. The use of HSI is effective in classifying different materials, because all materials absorb, reflect, or emit electromagnetic radiation when exposed to light [60]. Different materials produce distinct spectral information known as their spectral signature [7]. Therefore, hyperspectral imaging provides distinct information for identification of various materials.

HSI captures information in two spatial dimensions across hundreds of spectral channels [60]. Typically an HSI has small sampling intervals of 1nm to 15nm and covers the spectral range from the visible wavelength through the short-wave infrared wavelength [7]. A hyperspectral imager produces a highly dimensional data set, commonly referred to as a hypercube [60]. HSI data can provide detailed and valuable information when they are processed efficiently and effectively.

1.1.2 Feature Selection.

Hyperspectral imagers create data sets that are extremely large and difficult to process, making it necessary to reduce the data size for effective material identification. A typical spectrum from a spectroradiometer contains up to 2100 dimensions. High dimensional data are known to contain redundant features that interfere with the classification processes [36]. Extracting only the wavelengths that are crucial to accurate material identification is important with hyperspectral data.

A feature selection process extracts a meaningful set of features providing adequate information for identification of a material [42]. The work in this thesis focuses on determining a feature selection method and feature set that produce a 'fingerprint' capability of textiles.

1.2 Problem Statement

Recent advances in sensor technology has lead to in-depth research in skin and clothing detection in efforts to build a more accurate method for dismount detection [60]. However, current methods to detect clothing are inadequate for the use of uniquely identifying a dismount.

This thesis builds on an overlying research question: "Is it possible to uniquely identify or fingerprint textiles?"

In order to answer this question, a set of more specific questions must be answered: "What makes a textile unique?" "What features (wavelengths) from the spectral reflectance of textiles are necessary for identification?" "Is there a global set of features that can be used for all types of textiles?" "What kind of feature selection and classification methods are best fit for this data set?"

1.3 Scope and Limitations

This thesis focuses on fingerprinting different types of textiles. The goal for dismount identification is to be able to uniquely identify a person in a variety of environments. Textile identification will be an integral part of accurate and robust dismount detection.

Data used in this work are collected using a spectroradiometer with a contact probe, rather than using an actual HSI system. Because the contact probe minimizes outside noise and atmospheric effects, the data collected are treated as truth data. The assumption that these sets of truth data do not need to be corrected for noise, illumination variation or atmospheric effects is made. Feature selection and classification for uniquely identifying

textiles are accomplished using these truth data sets. All textile samples used in this thesis are one solid color without any stripes or patterns. The dye process or the thickness of the textiles are not considered as variables.

1.4 Methodology

Data are collected using a contact probe of a spectrometer, ASD FieldSpec ® 3 Hi-Res Spectroradiometer [34]. The spectroradiometer collects a material's spectral signature from the visible wavelengths to the short wave infrared wavelengths. All samples are represented as a continuous spectrum from 350nm to 2500nm. For each sample of fabric used, 100 samples are collected. Each sample is an average of 10 continuous instances of the material's spectral reflectance.

This research focuses on differentiating between textile materials of the same make. In order to accomplish this goal, the spectral data of different pristine fabrics are collected. Fabric swatches are machine washed to create separation from the pristine samples, taking measurements after washing. A different set of the same fabric swatches are washed and machined dried, also taking spectral data between cycles.

Several feature selection methods, including ReliefF, Bhattacharyya, fast correlation-based filter (FCBF), and Support Vector Machines with the aid of Recursive Feature Elimination (SVM-RFE), are performed on textile reflectance data collected. Classification methods including SVM, radial basis function (RBF) networks and multilayer perceptrons (MLP), are tested on the selected feature sets. The methodology is motivated from the fact that clothing articles are affected from the different wear and tear and maintenance techniques. Washing machines and dryers alter the spectral signatures of textiles, making it possible to uniquely identify a fabric of interest (FOI).

1.5 Results

When an FOI is evaluated by the first stage of the textile fingerprinting model, the color and the composition of the current FOI are classified and learned. With this knowledge, an ANN is trained specifically for identifying and discriminating against the similar colors and the composition of the FOI in question. This textile fingerprinting model is illustrated in Figure 1.2.

Each layer of the textile fingerprinting model is built with ANNs with MLPs. The relevant features are selected by SVM-RFE and FCBF. The classification for color and composition of textiles render a 100% accuracy when validated. The fingerprinting averaged accuracy with three FOIs is calculated to be 98%.

1.6 Overview

Chapter 2 is a review of important background concepts regarding clothing detection and research efforts made in this area. Related works accomplished in dismount detection, feature selection and classification using HSI data are presented. Feature selection and classification methods used to fingerprint clothing are described in Chapter 3. Chapter 4 reports the results of the data analysis accomplished for this thesis. Multiple feature selection and classification methods are compared to determine which methods work best. Chapter 5 summarizes the work done in this thesis and recommendations for future work.

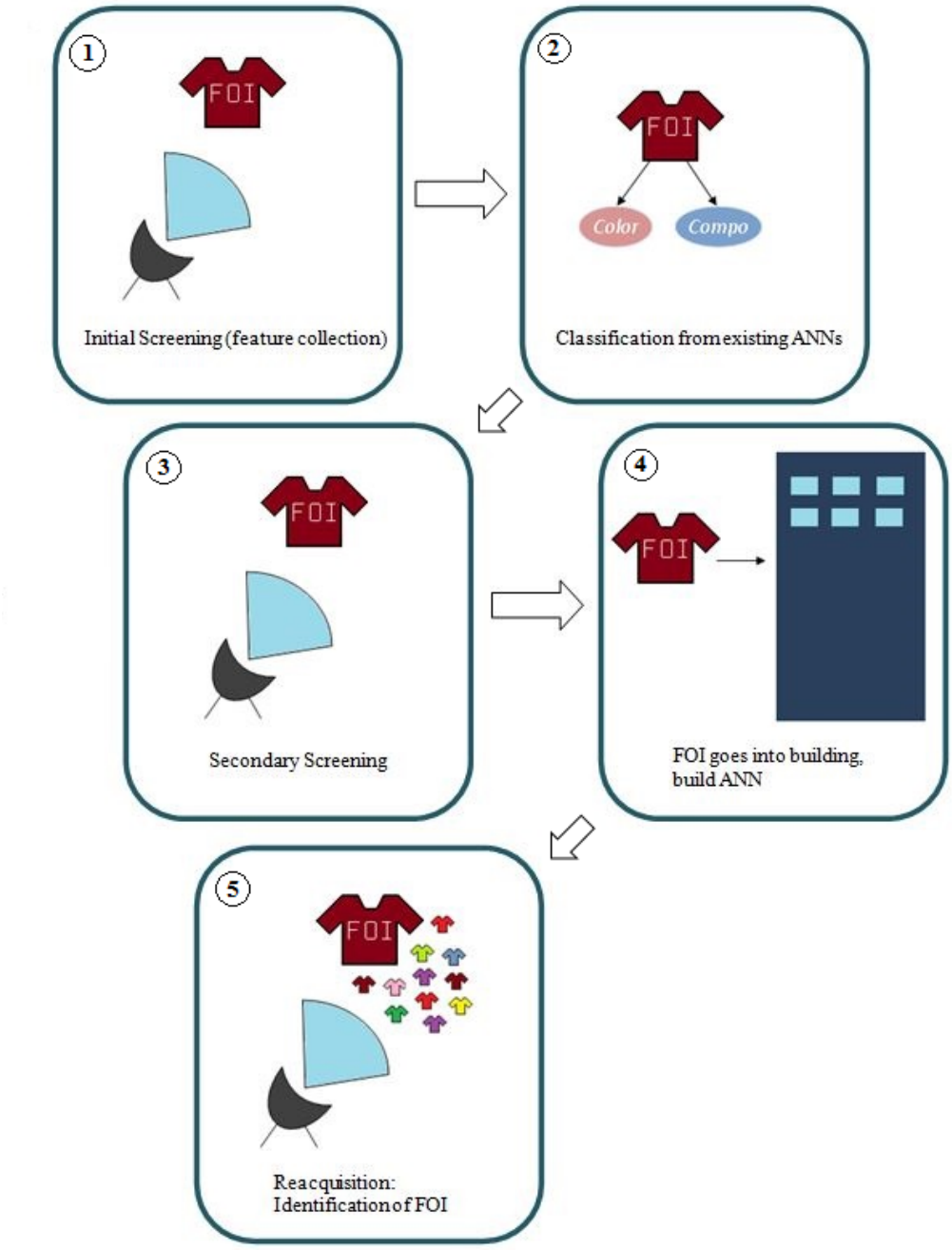


Figure 1.2: Visual diagram of the textile fingerprinting process. Initial screening collects features needed to classify the FOI into a specific color and composition. Secondary screening collects features according to the classification. An ANN is built with features collected in the secondary screening. The ANN is used for reacquisition/identification of the FOI.

II. Background

This chapter reviews research explaining certain concepts in textile fingerprinting and recent advances in this field of work. Central ideas in dismount detection, hyperspectral data classification, target detection, clothing detection, material identification and fingerprinting methods are presented. Background information on relevant concepts, algorithms, processes and procedures in hyperspectral data analysis are discussed.

Section 2.1 presents a broad view of dismount detection. Specific examples and explorations are presented in Section 2.1.1. Section 2.2 discusses hyperspectral imaging (HSI) and its applications. The basic theory of feature selection is reviewed in Section 2.3 with current feature selection methods in Section 2.3.1. Section 2.4 explains the recent work accomplished regarding clothing detection and classification. Characteristics of fabrics such as reflectance and transmittance are discussed in Section 2.5. Finally, in Section 2.6, other related works in material detection are reviewed.

2.1 Dismount Detection

Detection of humans, also known as dismount detection, has utilities in multiple applications, both civilian and military. Crowd monitoring, search and rescue, weapon targeting, industrial safety, security systems, and biometric identifications are a few of the most prominent applications of dismount detection. A variety of sensors can be used for dismount detection: electro-optical cameras, HSI, thermal infrared sensors and synthetic aperture radar (SAR) [57]. Dismount detection can be segmented into temporal, spatial, and spectral detection techniques [3]. Temporal techniques use SAR and the unique motion of the human frame; dismounts in motion are detected by their distinctive breathing and limb movements [57]. Spatial techniques involve searching for human

shaped geometric figures in a scene [30]. Dismounts are detected using a physiological dismount model of 12 different body parts [30]. To reduce the search space, spatial dismount detectors often search for areas that contain skin colored pixels [30]. Spectral dismount detection techniques exploit the known spectral signature of dismounts such as the unique reflectance of skin, hair, or clothing [3].

2.1.1 Related Works in Dismount Detection.

Rangaswamy *et al.* [57] present an approach to detect dismounts using synthetic aperture radar (SAR) and the unique signatures of dismounts present in SAR images. A study by Brooks [10] focuses on the full body detection approach using histograms of oriented gradients (HOG) features along with linear support vector machines (linSVM). A survey of the existing monocular dismount detection techniques is presented by Enzweiler *et al.* in [20].

Blagg [5] reports on a new detection approach that uses hyperspectral data in conjunction with thermal imaging to detect humans from other materials and the background. Thermal imaging exploits the temperature differential of dismounts to their surrounding. However, thermal imaging alone is not ideal since other heat sources, such as animals, may be present in the scene confusing the system. Blagg uses a matched filter algorithm which is performed on a spectral data set of a scene to search for clothing [5]. The results of the matched filter are thresholded and combined with the thermal data to give a final probability map for the scene under inspection [5].

An algorithm, denoted the Normalized Difference Skin Index (NDSI), uses the relationship between the spectral reflectance of skin at 1100nm and 1400nm for a computationally efficient skin detection [55]. NDSI is able to detect pixels of skin in a given scene without significant false alarms from objects that are skin-colored or contain water based liquid (vegetation and dirt) similar to the human skin [55]. This algorithm is

efficient at detection of skin, however, it is unable to uniquely identify a specific dismount from other dismounts.

Vehicle classification and recognition has also received a lot of attention during recent years. Roller *et al.* [58] uses a 3D generic vehicle model, parameterized to express different vehicles such as a sedan, station wagon, van or a bus. A simple sedan model and a probabilistic line feature grouping scheme are used for fast vehicle detection [41]. In [24], Duda *et al.* address the problem of vehicle matching and fingerprinting utilizing the edges of vehicles as features of discrimination.

2.2 Hyperspectral Imaging

HSI involves collecting and processing high resolution information across the electromagnetic spectrum [60]. Hyperspectral data typically include information from the visible spectral band to the shortwave-infrared band [60]. Hyperspectral data collection produces a continuous, highly dimensional data set of the spectral reflection [60]. The imaging spectroscopy principles, for remote sensing, are illustrated in Figure 2.1. Hyperspectral imaging is a passive analysis technique that can be used to obtain information from objects. This technique can be used to distinguish or identify materials based on their unique characteristics across the electromagnetic spectrum. HSI can detect signs of degradation, enhance the visibility of obscured features, or study the effect of environmental conditions on an object [49]. These 'fingerprints' are known as spectral signatures that enable identification of specific materials. Hyperspectral imaging is used in a variety of applications including chemical imaging in pharmaceutical research and industry, analysis of meat qualities, forensics, cosmetics and art [8, 18, 46, 56].

While HSI is highly informative, it requires excessive processing time and memory allocation to utilize because of its high dimensional nature. To reduce processing time, features that contain significant discriminating elements of a data set are selected [50].

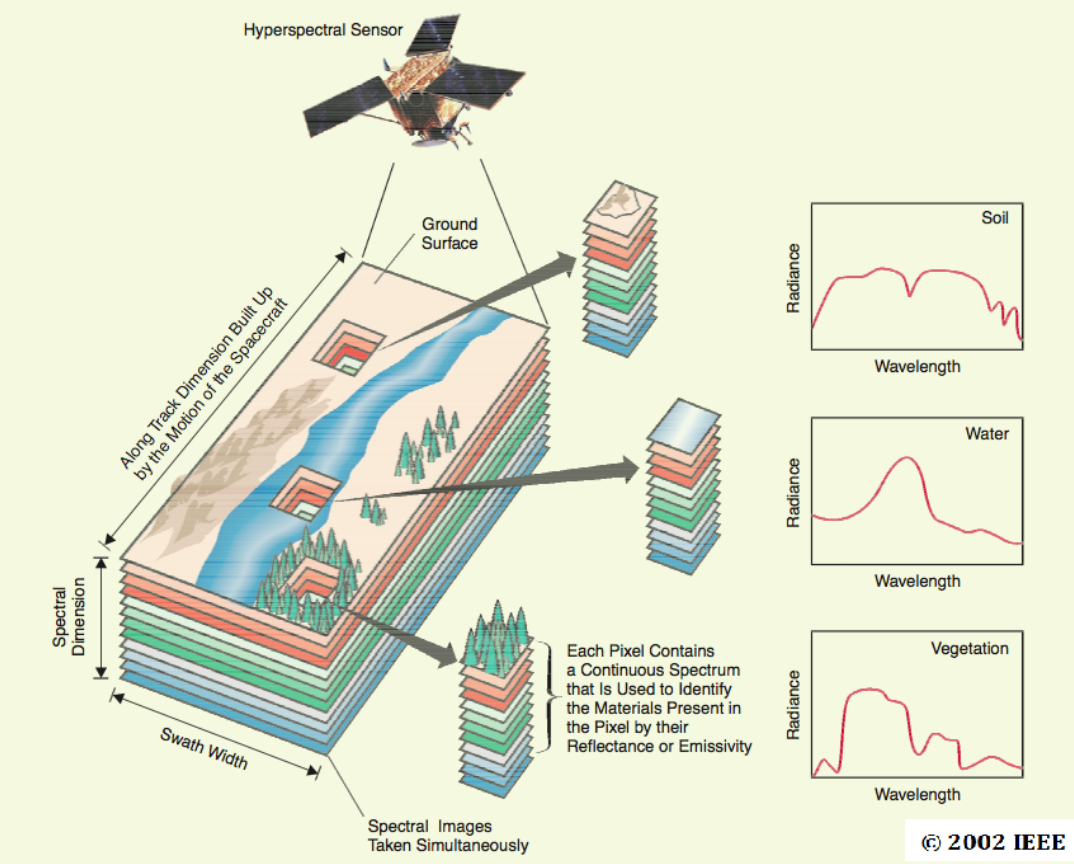


Figure 2.1: Principle of imaging spectroscopy [60]. Radiance from each spectral dimension is collected as a layer of the hypercube of the imaged scene. The three plots on the right hand side depict what the radiance of a single pixel of the image may show, depending on the material type.

2.3 Feature Selection Methods

The goal of a feature selection method is to preserve the classification capability, while minimizing the dimensionality of the data [14]. High-dimensional data sets often contain redundant and dependent features as well as missing data wasting computational resources as they provide no new information to the classification capability [60]. Therefore, it is

crucial to find the most appropriate features that contain significant information for a given classification problem [61].

Dash and Liu [17] classify feature selection into four different definitions: idealized, classical, improving and approximating. The idealized method finds the smallest available feature set that is sufficient to classify or detect given targets [42]. The classical approach selects an optimal subset of m features from an n-feature set [17]. Improving prediction accuracy method focuses on selecting a small subset of features and decreasing the size of that set without decreasing the classification accuracy [45]. The approximating original class distribution approach selects the smallest possible feature set, resulting in a class distribution that is as close as possible to the original class distribution [45].

For a feature set of size N, there will be 2^N subsets of the feature set available [17]. Looking for the best subset using every combination of features may be exhaustive and too costly even with a medium sized feature set. Dash and Liu [17] state that all feature selection methods have four basic steps: generation procedure, evaluation function, stopping function and validation procedure.

The generation procedure produces subsets of features for evaluation using three different search methods: complete, heuristic and random [17]. A complete search exhaustively searches the data space in order to find the optimal subset [17]. A heuristic search uses a function in order to determine the cost or 'goodness' of a feature [17]. The cost is calculated at every iteration and determines if a feature is selected or rejected. A random search randomly evaluates a feature and uses a function to determine if that feature should be selected, similar to the heuristic search [17].

An evaluation function measures the 'goodness' of a subset selected by the generation procedure [17]. An evaluation function will measure the discriminating ability of a feature or a subset to distinguish the different classes [17]. The evaluation function retains the best subset according to the goodness measurements. The authors Dash and Liu [17]

classify the evaluation functions into five categories: distance, information (or uncertainty), dependence, consistency and classifier error rate. Distance measures use the separability of features and is a metric for describing how 'far away' or 'close to' elements are from each other. A widely used example of a distance measure would be the Euclidean distance measure. Information measures determine the information gain from a feature; for example, entropy, which is a measure of the uncertainty in a random variable [17]. Dependence measures makes use of the correlation between a feature and a specific class [17]. The correlation between classes can be utilized to determine which classes are close, or related to each other. The correlation between features indicate the degree of redundancy of the features. Consistency measures determine the minimally sized subset satisfying the acceptable inconsistency rate set by the user [17]. The classifier is the evaluation function for the classifier error rate measure, resulting in high accuracy levels as well as high computational costs [17].

The stopping criterion keeps the feature selection process from running exhaustively or redundantly through the groups of subsets [17]. A stopping criterion can be a preset number of features/iterations, or the goodness value of a subset based on the addition/deletion of other features [17]. The validation procedure is not part of the feature selection, but is necessary for any feature selection method. It tests the validity of the selected subset by comparing results with previous findings by other feature selection methods, performing different tests, or applying the subset to a real world data set [17].

Langley [49] divides feature selection methods into three groups based on the feature selection and evaluation strategy: embedded, filter and wrapper. The embedded method determines a features' goodness during the feature selection process [49]. The feature selection process is embedded within the basic induction algorithm [49]. The filter method introduces a separate preprocessing step based on the general characteristics of the training set before the basic induction step that will select features [37]. The wrapper

method uses the inductive algorithm as the evaluation function [49]. Feature selection methods that use the classifier error rate as an evaluation measure are considered wrapper methods [17]. This method is based on the argument that the induction method will provide a better estimate of accuracy than a separate measure that might have a different inductive bias [49].

2.3.1 Current Methods for Feature Selection.

A popular method of feature selection is the Principal Component Analysis (PCA). PCA is a dimension reduction process that uses an orthogonal transformation to convert a set of correlated variables or features into a lower dimensional data set of uncorrelated features [38]. Principal components are computed by using the eigenvectors and eigenvalues of the covariance matrix of the data set as represented in Algorithm 1 [15].

Algorithm 1: Computing PCA of a given data set [15].

Input: $x = [x_1, x_2, ...x_n]$: training examples

m: number of principal components to keep

Output: $PC = []$: principal components

1 $x_z = x - \text{mean}(x)$: zero mean the data

2 $\Sigma = \text{cov}(x_z)$: calculate covariance matrix

3 calculate eigenvalues, eigenvectors of Σ

4 rank eigenvalues (largest to smallest)

5 $PC = m$ largest eigenvalue/vector pair

6 **return** principal components PC

7 Project data onto selected principal components

Only a small number of eigenvectors are necessary to reduce the dimension of the data while maintaining classification accuracy. This group of chosen eigenvectors form a new

basis for the data set [15]. The data are projected by the selected eigenvectors into a feature space.

The eigenvectors characterize the direction of the variance of the data whereas the eigenvalues are the characteristic value of their corresponding eigenvectors [1]. The larger the eigenvalue, the more important the corresponding eigenvector or principal component [15]. Overall, PCA is able to condense a large data set with n dimensions into a data set of m dimensions where $m < n$, maximizing the covariance and reducing the redundancy of a particular data set [1].

Although PCA is a useful tool for dimension reduction and data compression, it is not typically suitable for feature extraction when attempting target detection or classification [13]. Although easy to implement, PCA is inept at extracting discriminating features from certain data sets as the higher order components do not always retain the discriminatory data of the original feature set [13].

Learning vector quantization (LVQ) is a neural learning algorithm that uses prototype vectors to determine a decision boundary for classification [44]. Prototype vectors are the trained quantities in an LVQ that learn a given representation of the assigned class [4]. The LVQ algorithm iterates over the training data and updates the prototype's position while defining class boundaries [4]. Generalized relevance learning vector quantization (GRLVQ) is a variant of LVQ using gradient-descent algorithms to provide feature ranking information based on a feature's discriminatory capability [26]. Supervised learning algorithms such as LVQs are desirable because they are robust to noisy samples and incomplete data [4].

The Bhattacharyya coefficient method is a feature selection method that utilizes the probability distributions of the features [15]. The Bhattacharyya coefficient measures the area of overlap between distributions of two different features. The Bhatacharyya coefficient measures the separability between histograms. The coefficient is calculated as:

$$B_n = \sum_{k=1}^{K} \sqrt{P_X(k)P_Y(k)}, \tag{2.1}$$

where B_n is the Bhattacharyya coefficient for feature n and its instance k, $P_X(k)$ is the probability of k for probability distribution function (PDF) of X, $P_Y(k)$ is the probability of k for PDF Y and K is the number of bins in both histograms [16].

The Bhattacharyya coefficient (B_n) has a maximum value of 1 when the two distributions being compared completely overlap [15]. The coefficient becomes smaller as the separation between the two distributions get larger. The coefficient has a 0 value when there is no overlap of the distributions [15]. A visual representation of the coefficients can be seen in Figure 2.2. Three sets of distributions are shown: one where the two distributions nearly overlap completely, one where they overlap half way, and one with no overlap. The corresponding Bhattacharyya coefficient can be seen alongside each graph. The Bhattacharyya coefficient is calculated for every feature of the data set. Features with smaller coefficients are selected when reducing the dimensionality of a given data set and still preserve class discrimination.

Relief is a feature ranking method that uses a distance measure to determine the weights for every feature [15]. The Relief algorithm was motivated by nearest-neighbors, aligning features of instances of the same class and differentiating from features of the other class [42]. Relief is a two class binary feature selection method. ReliefF is adapted for multiclass problems as well as noisy and incomplete data sets [42].

The Relief algorithm assigns a weight to each feature which denotes the relevance of a feature regarding the discriminability of the data [15]. Reilef randomly selects a sample of instances and finds its nearest hit and nearest miss instances based on a selected distance measure [15]. The nearest hit is the instance that has the minimum distance to the selected instance of the same class as the selected instance. The nearest miss is the instance with the minimum distance to the selected instance of a different class as the selected

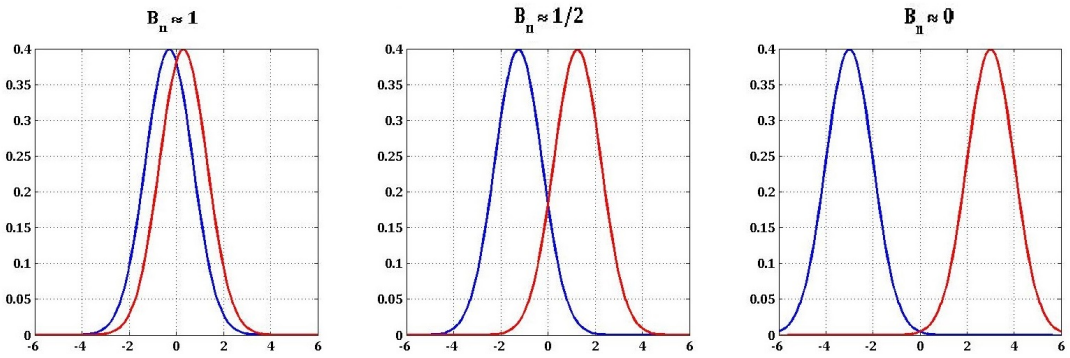


Figure 2.2: Three different distribution sets with corresponding Bhattacharyya coefficients. The red and blue graphs are the PDFs of the two distributions being compared. The left most graph has the most overlap with $B_n \approx 1$, the middle graph has half overlap with $B_n \approx 1/2$, and the last graph has no overlap with $B_n \approx 0$.

instance [15]. The feature weights are initialized to zero initially and are updated by evaluating the nearest hit and nearest miss distances from the sample instance [15]. Relief works well for noisy and moderately correlated feature sets both for nominal and continuous data. Relief, however, does not filter out redundant or highly correlated features, and often generates non-optimal feature sets in the presence of redundant features [43].

Support Vector Machine (SVM) is a supervised learning algorithm for classification and regression analysis. SVMs determine a decision boundary with maximum margins between the data sets of the two classes [25]. SVMs are discussed in further detail in Section 2.4.1.

A feature ranking method proposed by Guyon, *et al.* utilizes SVMs based on Recursive Feature Elimination (RFE) [25]. RFE is a backwards feature elimination procedure used for feature ranking [25]:

1. Train the classifier (SVM - Algorithm 4)

2. Compute ranking criterion for all features, $(w_i)^2$

3. Remove feature with smallest criterion

The SVM-RFE feature selection produces a feature ranking based on the weights of the SVM classifier [25]. The weights of the decision function of an SVM are only a small subset of the training samples, referred to as "support vectors" [29]. The weight vector is determined by a linear combination of the training samples and the non-zero weights are regarded as the support vectors [15]. The pseudocode for SVM-RFE is shown in Algorithm 2.

Algorithm 2: SVM-RFE proposed by Guyon *et al.* [25] ranks features from best to worst using the recursive feature elimination method.

Input: $\boldsymbol{X_0} = [\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_k}, ...\boldsymbol{x_l}]^T$: Training examples

$\boldsymbol{y} = [y_1, y_2, ...y_k, ...y_l]^T$: Class labels

initialize: $\boldsymbol{s} = [1, 2, ...n]$: Subset of surviving features

Output: $\boldsymbol{r} = []$: Feature ranked list

while $s \neq NULL$ **do**

  $X = X_0(:, s)$: restrict training examples to good feature indices

  $\alpha =$ SVM-train(X, y): train the classifier

  $\boldsymbol{w} = \sum_k \alpha_k y_k \boldsymbol{x_k}$: compute the weight vector of dimension length

  $c_i = (w_i)^2, \forall i$: compute the ranking criteria

  $f = argmin(\boldsymbol{c})$: find the feature with smallest ranking criterion

  $\boldsymbol{r} = [\boldsymbol{s}(f), \boldsymbol{r}]$: update feature ranked list

  $\boldsymbol{s} = \boldsymbol{s}(1 : f - 1, f + 1 : length(\boldsymbol{s}))$: eliminate feature with smallest ranking

return *Feature ranked list r*

The Non-correlated Aided Simulated Annealing Feature Selection - Integrated Distribution Function (NASAFS-IDF) feature selection method uses a distributed spacing function with the discrimination capability of the feature set as a heuristic for the collection of a more robust and accurate feature set [14]. NASAFS-IDF is an optimized feature selection method that maximizes the given heuristic by utilizing a simulated annealing search [14]. The simulated annealing approach increases the chance of evading locally optimal solutions [14].

NASAFS-IDF involves dividing the data into a number of bins, determining the cross-covariance threshold, selecting a feature set randomly, evaluating the heuristic, then replacing a feature in the set with another random pick from the remaining features and re-evaluating the heuristic [14]. This process is repeated until a stopping criteria is met, creating a non-correlated feature set [14].

A fast correlation-based filter (FCBF) solution is proposed by Yu and Liu which uses information gain to analyze feature redundancy to select features that are relevant and not redundant [51]. The FCBF algorithm uses symmetrical uncertainty (SU) to determine the relevance of each feature. SU is defined as [51]:

$$SU(x, y) = 2\left[\frac{IG(x|y)}{H(x) + H(y)}\right], \tag{2.2}$$

where IG is information gain, H is entropy, x and y are features. A feature becomes relevant if the SU value is greater than a user defined threshold. The threshold is denoted δ and the correlation between a class C and a feature f_i is denoted as $SU_{i,c}$ [51]. SU values are between 0 and 1 where the SU value of 0 indicates that two features are independent whereas an SU value of 1 indicates that the value of X can be completely predicted by Y [51].

Heuristics are used to determine redundancy of the feature. The feature with the highest SU value is deemed predominant. A feature is removed if it is found to be redundant to a more predominant feature. This process is continued until there are no

features to be removed, ultimately producing the optimal feature subset of the given data [51]. The pseudocode for FCBF is shown in Algorithm 3.

Algorithm 3: FCBF finds the optimal feature subset data set using the symmetrical uncertainty value and heuristics [51].

Input: $S(f_1, f_2, ..., f_N, C)$: labeled training data set

δ: relevance threshold (user defined)

Output: S_{best}: selected features

```
for i = 1 to N do
    SU_temp = SU_{i,c} for f_i
    if SU_temp ≥ δ then
        add f_i to S_list
Sort S_list in descending order
f_a = firstElement(S_list)
while f_a ≠NULL do
    f_b = nextElement(S_list)
    while f_b ≠NULL do
        if SU_{a,b} ≥ SU_{b,c} then
            remove F_b from S_list
        f_b = nextElement(S_list)
    f_a = nextElement(S_list)
S_best = S_list
return S_best
```

2.4 Classification Methods

Classification is the process of assigning a label to an observation, whereas detection is the process of identifying the existence of a condition. Detection can be considered as a two-class classification problem: target exists or does not exist [60]. Detection can also be referred to as binary classification [60]. In detection applications, *a-priori* information about the desired target is necessary [39]. However, in most applications, the target class occupies only a small population in the scene, and the background is referred to as everything except the target in a given scene [35].

The goal of target detection is to maximize the probability of detection while minimizing the false alarm rate [35]. Many algorithms use a constant false alarm rate (CFAR) detector to keep the false alarm rate constant at a specific desired level [35]. A CFAR algorithm provides detection thresholds that are robust to noise and background variations [60].

Detection algorithms generally compare an observation to a known spectral signature, an expected distribution of signatures or a set of detection rules to classify a given observation [23]. Detection algorithms such as anomaly detectors (AD) that do not require prior knowledge determine regions of interest based on the statistically distinct spectrum of one region from the background [62]. This detection technique only uses the difference of spectral signatures, therefore, it does not require *a-priori* knowledge to perform detection [40]. Examples of AD models are the local Gaussian model, the global Gaussian mixture model and the global linear mixture model [62]. Other detection methods use supervised learning and *a-priori* knowledge of specific signatures or given materials [61].

Spectral detection can be divided into two categories: those that utilize reflectance-based detection and those that utilize radiance-based detection [3]. Reflectance-based detection involves atmospheric compensation as one of its preprocessing steps whereas radiance-based detection skips the atmospheric compensation

and performs noise whitening and normalization during its preprocessing for the detection algorithm [3]. Hyperspectral detection algorithms that search for a specific target signature work the best [3]. These types of detection algorithms compare the spectral signatures from an observation to a specific target signature, giving a measurement based on their similarity, and comparing the measurements to a threshold to determine the target's classification [3]. A binary classifier classifies each item as a target or a non-target providing the probability of detection and the probability of false alarm [60]. Setting different thresholds on the measurement provides different probabilities of detection and false alarm [3].

2.4.1 Current Classification Methods.

The No Free Lunch Theorem in [19] states that there is no overall best classifier. If one classifier outperforms another in a particular set of circumstances, it is because of the fit to that particular pattern recognition problem, not because of the general superiority of the classifier. The decision on which algorithm to use is based on aspects such as prior information, amount of training data, cost or reward functions and data distribution [19].

ANNs, such as, radial basis function networks, multilayer perceptrons, and support vector machines, can be utilized as a classifier when trained with a training set representative of the data set to be classified [29]. A typical ANN processes the training set through the network while calculating the error between the actual and desired outputs, updating the network until a desired error threshold is met.

A RBF network typically has three layers, to include the input layer, one hidden layer and an output layer [29]. The input layer consists of input nodes that connect the network to the environment. The number of input nodes used corresponds to the number of features or dimensions of the inputs [29]. The second layer, referred to as the hidden layer, is made up of nodes that apply a nonlinear transformation to the input, transforming it to

the feature space [29]. The output layer is a linear combination of the transformed inputs and neuron parameters [29].

A network of multilayer perceptron contains neurons with differentiable activation functions [29]. The network is made up of one or more hidden layers. The number of hidden layers is determined by the user. The multilayer perceptron network generally has a high degree of connectivity [29]. An architectural graph of a multilayer perceptron can be seen in Figure 2.3.

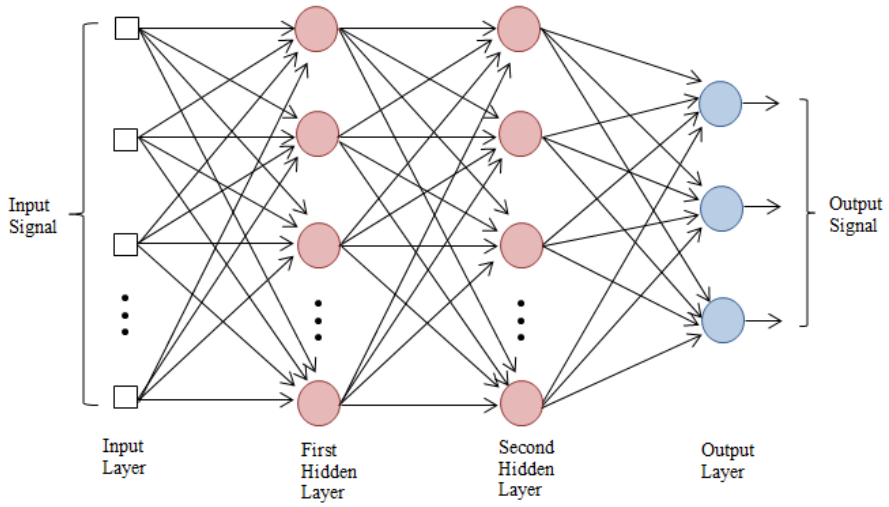


Figure 2.3: Architectural graph of a multilayer perceptron with two hidden layers. The input signal comes in from the left to the square nodes to the input layer, pink circular nodes are the hidden nodes that the information from the inputs get passed to. Blue nodes depict the output nodes. The arrows show the flow of the forward pass, showing that this is a fully connected perceptron. The back-propagation is not shown on this graph for simplicity.

Training of a back-propagation multilayer perceptron can be segmented to two major phases [29]. The first phase is the forward pass, where an input is passed through the

network and is acted upon by the network to produce a specific output. The second phase is the back propagation, where the error is calculated by comparing the output and the desired response, and is propagated back through all layers of the network, making adjustments to the weights as necessary [29].

A support vector machine (SVM) is a binary classification method that constructs an optimal hyperplane as a decision boundary [29]. SVMs are designed as a feed-forward network with a single hidden layer containing nonlinear properties [16]. The hyperplane is built to maximize the margin of separation between the two classes [29]. The optimality of a SVM is accomplished using convex optimization [29]. The support vectors used in SVMs are a subset of input data points extracted from the learning algorithm also referred to as a kernel method [29]. Algorithm 4 shows the training executed in an SVM.

Algorithm 4: Training weights of an SVM given a data set and corresponding label [25]

Input: $\boldsymbol{x_1}, \boldsymbol{x_2}, ...\boldsymbol{x_k}, ...\boldsymbol{x_l}$: Training samples

$y_1, y_2, ...y_k, ...y_l$: Class labels

1 minimize over α_k:

2 $J = (1/2) \sum_{hk} y_h y_k \alpha_h \alpha_k (\boldsymbol{x_h} \cdot \boldsymbol{x_k} + \lambda \delta_{hk}) - \sum_k \alpha_k$

3 subject to:

4 $0 \leq \alpha_k \leq C$ and $\sum_k \alpha_k y_k = 0$

Output: parameters α_k

In Algorithm 4, the summations are accomplished over all training samples x_k. The Kronecker symbol represented with, δ_{hk}, is either 1 if $h = k$ or 0 otherwise. Positive constants λ and C are user defined to ensure convergence even when a problem is non-linearly separable. A small λ (of the order of 10^{-14}) is used to ensure numerical stability. After the training in Algorithm 4, a decision function $D(x)$ is made [29]:

$$D(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b, \tag{2.3}$$

with

$$\boldsymbol{w} = \sum_{k} \alpha_k y_k \boldsymbol{x_k}, \tag{2.4}$$

and

$$b = \langle y_k - \boldsymbol{w} \cdot \boldsymbol{x_k} \rangle. \tag{2.5}$$

The weight vector w is a linear combination of training samples where most weights α_k are zero [25]. The bias value b is an average over marginal support vectors, which are greater than zero and smaller than a set parameter C [25].

2.5 Clothing Detection and Classification

Clark [14] presents a feature selection method, NASAFS-IDF, described in section 2.3.1. To test the algorithm, he uses NASAFS-IDF versus other feature selection methods, to discriminate 12 different classes of textiles based on their spectral signature collected across the 900nm - 2450nm wavelength range under several noise realizations.

Haran collects several diffuse reflectance measurements of polyester and cotton and characterizes the spectral features in the short-wave infrared spectrum (wavelengths between 0.9 and 2.5 microns) [28]. He identifies a set of unique spectral fingerprints in the SWIR region that allows him to distinguish cotton and polyester fabrics regardless of their visible color or texture [28].

Additionally, there has been research for developing fabrics with reduced reflectance signatures in the visible and near infra-red spectral bands in [22]. Frankel *et al.* is able to modify the NIR signature of the base fabric by the inclusion of nano and micro particle additives to the original fabric in order to enhance concealment [22].

Another emerging area exploiting hyperspectral data of textiles is the detection of defects. The quality control of fabrics is a vital step for the modern textile industry, as defects can have a large impact on grading and costs of the final product [47]. Zhou *et al.* [66] presents a fabric detection scheme using sparse dictionary reconstruction. In this study, fabric defects are regarded as local anomalies against a homogeneous texture background [66]. Kumar *et al.* [48] presents an approach for automated inspection of textured materials for detection of defects using Gabor wavelet features.

2.6 Fabric Reflectance and Transmittance

Various manufacturing processes exist for creating textiles. Textiles may be production-line manufactured where thousands at a time are made of the same color and material type. Textiles could be manufactured by weaving threads in a specific pattern. Other methods include knitting, matting, and compression of fibers into a piece of fabric. Figure 2.4 shows an example of the reflectance of cotton and polyester fabric samples across the VIS-NIR-SWIR spectrum.

The study of fabric reflectance can be traced to the 1960s when a study using an interferometer spectrometer was conducted to characterize fabrics by determining the materials' surface emissivity [54]. The different types of textiles were imaged at in the 1 um to 15 um wavelength region. Terrain features and ambient conditions in the field had a large effect on any measurement taken, making it difficult to correlate with the laboratory emission studies [54].

A more recent study by Herweg *et al.* [32] incorporates the knowledge that textiles are dense and opaque at visible wavelengths to the naked eye yet porous enough to allow sensors to detect a significant amount of energy. This property of clothing allows the spectral content of backing material to be reflected through the fabric [32]. Herweg *et al.* [32] use three types of fabrics as targets on asphalt and grass, with illumination

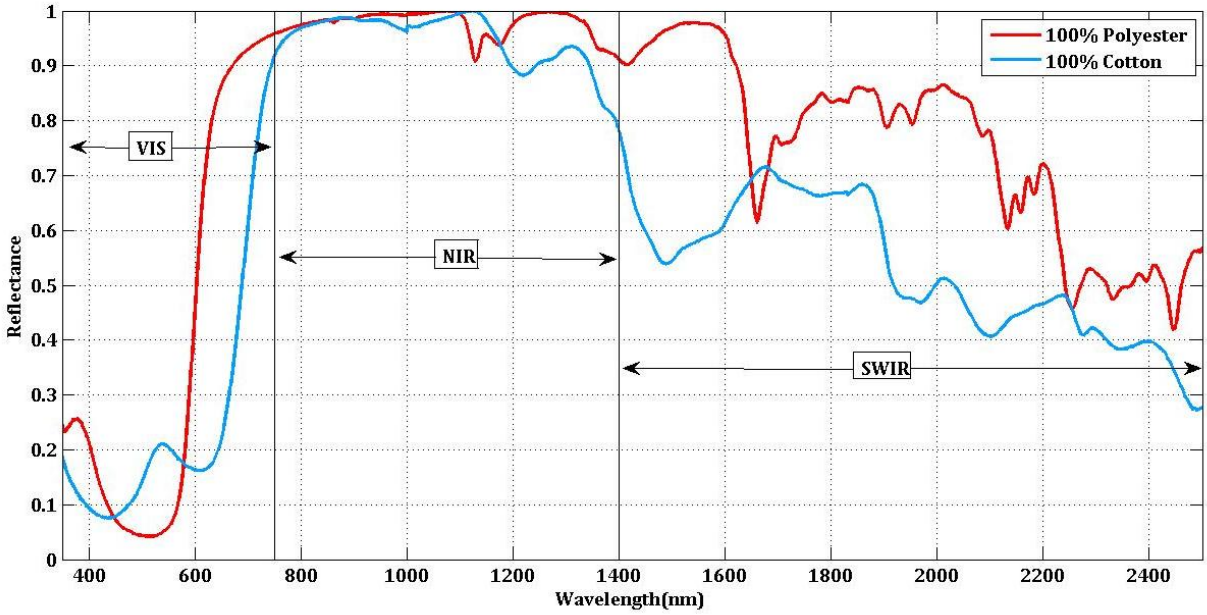


Figure 2.4: Normalized reflectance of two sample fabrics: cotton in blue and polyester in red. Includes normalized reflectance of the VIS through SWIR bands of both fabrics. VIS, NIR, and SWIR regions are depicted by the double sided arrows.

variations to examine the variability of fabric's target signature on different types of backings. Spectral measurements of the fabrics against different backings are collected to show the variation of these measurements based on the different backing [32].

2.7 Other Related Works in Material Recognition and Fingerprints

Material recognition is a field of study highly explored in computer vision. Liu *et al.* in [50] explores the problem of characterizing materials such as glass, metal, fabric, plastic or wood from a single image of a surface. The authors use a set of features that capture various aspects of the material's appearance and propose an augmented Latent Dirichlet Allocation (aLDA) model to combine these features [50].

The texture of materials is used for recognizing classes within remote sensing where the goal is to distinguish between background environments such as water, forest, and urban areas [12]. Gabor filters were used in a comparative study of several texture features for remote sensing data [59]. In [65], Zhang *et al.* use wavelet-based methods such as the hidden Markov model (HMM) to classify soil textures.

A study by Zhang *et al.* [64] combines multiple features such as texture and shape features as well as hyperspectral features for image classification purposes. Each feature used provides a particular contribution to the overall representation determined by optimizing the weights in the objective function [64]. Zhang *et al.* is able to successfully achieve a low-dimensional representation for an accurate and effective classification [64].

A fingerprint of a material, similar to a human's fingerprints, is unique to its corresponding material. Bruce *et al.* [11] proposes using a derivative operation in the analysis of hyperspectral signatures. A derivative operation measures the changes of a function as the input changes [11]. When it is applied to a hyperspectral curve, the characteristic shape of the contours gives a spectral fingerprint as the locations of the inflection points and their slopes can provide important information about the spectral features [11].

2.8 Summary

This chapter reviews studies done in the field of dismount detection and hyperspectral imaging. An overview of the fundamentals of hyperspectral imaging are presented as well as some of its applications. The principal theories of feature selection are discussed and a few common methodologies are presented such as PCA and GRLVQ. Recent efforts in the clothing detection and classification area are explored to find what has already been accomplished in this particular field. Studies that engage material recognition and fingerprinting are also reviewed. The background knowledge and literature review

determine that an optimal method to fingerprinting fabrics or textiles using feature selection and classification has not yet been achieved.

III. Methodology

Recent advances in sensor technology and hyperspectral imagery have made collecting spectrally high resolution radiance of materials throughout the electromagnetic spectrum possible [60]. The hyperspectral imaging (HSI) capability presents the possibility of a variety of material identification and characterization experiments [8, 24]. Researchers are able to distinguish among different materials more easily due to each material's unique spectral signature. These characteristics associated with hyperspectral imaging make HSI an essential tool for dismount detection.

Hyperspectral data of a person's skin and clothing are currently being studied in order to make current dismount detection methods more robust. Past research efforts have discovered that different textiles can be distinguished from each other using their spectral signatures [14]. The ability to characterize textiles is highly useful for dismount detection purposes.

This chapter introduces the methodology used to uniquely identify textiles using their spectral signatures. In section 3.1, the type of data used is introduced along with how the data are obtained. The process in which the data sets are analyzed is explained and justified in section 3.2.

3.1 Data Collection

Spectral reflection measurements are collected using the ASD FieldSpec ® 3 Hi-Res Spectroradiometer [34]. This instrument has a spectral range from 350nm to 2500nm which includes the visible (VIS), near infrared (NIR), and short-wave infrared (SWIR) spectral bands. It has a 1.5m fiber optic cable and provides 2151 channels or features for every measurement. Each measurement is given as the average of 10 instances.

A contact probe is used for collecting spectral data of textile samples. The contact probe is equipped with a broad-band halogen internal illumination source to emulate natural sunlight. The probe makes complete contact with the samples, minimizing noise and atmospheric effects in the measurements.

The spectroradiometer is calibrated before taking measurements with the data acquisition software supplied by ASD Inc [34]. The optimization step calibrates the spectroradiometer for the specific light source used for a data collect. The halogen light source within the contact probe is used in all data collections for this thesis. The white reflectance calibration collects absolute reflectance across 350nm - 2500nm using a spectralon panel, to produce reflectance values between 0 and 1.

Due to the porous nature of textiles, a fraction of the light source will penetrate through the existing pores of the sample and not reflect back [31]. In order to minimize this occurrence, the textile samples are backed with themselves by folding the material in half three times. A black reflectance panel is placed behind the folded textiles to standardize collection procedures. The laboratory setup of the data collection is shown in Figure 3.1.

Figure 3.1: Laboratory setup of reflectance data collection. The green, 100% cotton shirt sample, contact probe, and black calibration panel used are labeled.

Figure 3.2 shows an example of a textile reflectance measurement collected by the spectroradiometer. Each line color represents an instance of the reflectance of green, 100% cotton. These measurements are collected in 5 separate collection cycles, in 5 different regions of the textile. Twenty instances are collected in each region, totaling 100 instances per textile. The reflections from different regions are used to prevent over-fitting to one specific region of a textile during data analysis.

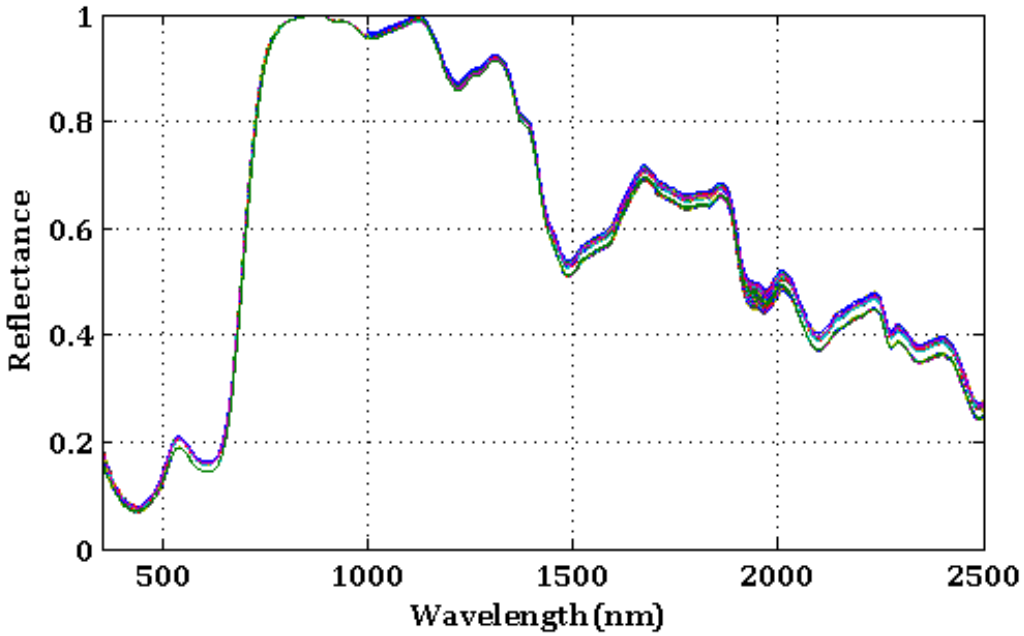


Figure 3.2: 100 Instances of Cotton Reflectance from ASD FieldSpec ® 3 Hi-Res Spectroradiometer using a contact probe, with a black reflectance panel as backing. Each instance is the average of 10 consecutive samples.

The data set used in this thesis consists of 8 different types and blends of textiles, as well as paint chip samples of 32 different colors. The data sets can be separated into three types (color, composition, and uniqueness). The paint chips used for obtaining the color data set can be seen in Figure 3.3. The colors consists of 8 different color schemes (red,

orange, yellow, green, turquoise, blue, indigo, and purple) each with four different shadings. The shading is dependent on the amount of gray that is added to the original color [9]. The textile samples used are listed in Table 3.1 with their matching colors. The composition data set consists of the reflection data of the pristine textile. In order to create separation between the same makes of textiles, a group is kept in its pristine state, whereas the other groups are washed and dried using different methods. A Kenmore® high efficiency washing machine is used on its warm cycle with a Tide® original laundry detergent. A Kenmore® standard dryer is used on its medium heat, 40 minute tumble dry cycle. The uniqueness data set is comprised of these modified textiles.

Figure 3.3: Color chip samples used in feature selection for visible spectrum. From left to right, the color schemes are: purple, indigo, blue, turquoise, green, yellow, orange, and red. Each color scheme has four shadings, separated by horizontal white lines.

After each data collection, a variety of feature selection and classification problems are conducted. One problem addresses the classification of different textile types such as cotton, polyester, and denim. Another problem focuses on the classification of washed and pristine versions of the same textile (uniqueness). A separate feature selection method is employed for each classification problem presented.

Table 3.1: List of textile materials and colors

Type	Color
cotton 100%	
polyseter 100%	
cotton 100% (denim)	
nylon 100%	
cotton 17% polyester 83%	
cotton 30% polyester 70%	
cotton 61% polyester 34% spandex 5%	
cotton 88% nylon 12%	

Each data set is divided into training and validation sets. The training set is used to train an ANN for a classification problem, using K-folds cross validation. The validation data set is set aside for the purpose of testing the ANN model produced. The two sets are created by randomly shuffling the original data set and dividing the instances into two buckets of the desired percentage. The division of each of the data sets can be seen in Table 3.2.

Table 3.2: Separation of training and validation sets for each data set.

Type	**Training** (%)	**Validation** (%)
Color	15	85
Composition	50	50
Uniqueness	50	50

3.2 Data Analysis

Data analysis consists of three major steps to include preprocessing, feature selection, and classification. The bulk of the feature selection and classification are completed using Weka (Waikato Environment for Knowledge Analysis), which is a machine learning software developed at the University of Waikato, New Zealand [52].

Weka supports numerous data processing tasks such as data clustering, classification, regression, feature selection and visualization. The feature selection methods available in Weka include filter and wrapper methods that uses a variety of search methods such as: best first search, exhaustive search, genetic, random and scatter search. Weka provides classifiers such as Naive Bayes, SVM, artificial neural network with multi-layer perceptrons, radial basis function networks, and decision trees.

3.2.1 Preprocessing.

Each sample is normalized to ensure that all data samples are compared on an equal baseline. Normalizing the data is necessary to prevent unwanted biases during the feature selection and classification process. The data are normalized using the following equation [15]:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{3.1}$$

where x is the original data, x' is the normalized data, $\min(x)$ and $\max(x)$ are the minimum and maximum values of the data sample.

Preprocessing of the data, to remove outliers and for normalization, is accomplished in Matlab ®. Matlab ® is also used to format the data sets into a csv file for the ease of data analysis in Weka.

3.2.2 Feature Selection.

The spectral reflection of 32 different colors samples are used for feature selection in the visible spectrum. The wavelengths used for feature selection in the VIS spectrum are 350nm to 750nm. The spectral reflection data of the pristine textile samples are used in

feature selection for material (composition) classification. Fabric color is classified separately, therefore, only the reflection data of the NIR and SWIR bands are used in efforts to classify material composition.

Both the VIS and NIR/SWIR data sets are processed with the SVM-RFE feature selection method discussed in Section 2.3.1. SVM-RFE is designed to rank all existing features from best to worst [25]. SVM-RFE is implemented to the data sets following the pseudocode shown in Algorithm 2. Each training data set is normalized, given corresponding class labels, and processed stepping through every feature recursively, in Weka. The input to the algorithm is a data set that contains f features with labeled classes. The output of the algorithm is a ranked feature list as processed by RFE. The top 4 features of the VIS spectrum and the top 5 features of the NIR/SWIR spectrum are selected from each ranked feature list. The specific results of these feature selections are discussed in further detail in Chapter 4.

Color and composition classification of textiles provides discrimination between colors and types of textiles. However, these features alone are inadequate for fingerprinting of textiles. A third set of features is selected for each textile class to uniquely distinguish between textiles of the same composition. A universal feature set for discerning color, composition, and uniqueness is difficult to obtain, therefore, the feature selection process is done separately in stages for each textile composition. Since feature selection is done for each composition, the fast correlation-based filter method (FCBF) [51] described in Section 2.3.1 is used instead of the computationally expensive SVM-RFE. FCBF is a filter-type method which identifies relevant features as well as redundant features without pairwise correlation analysis between all relevant features [51]. The FCBF feature selection method as shown in Algorithm 3 is implemented on each data set of as single textile composition with its different versions (washed, dried), in Matlab ®. The input of the FCBF algorithm is a data set of f features with labeled classes and a predefined

threshold, δ. The algorithm outputs an optimal feature set after eliminating redundant and irrelevant features.

3.2.3 Classification.

An artificial neural network (ANN), made up of multilayer perceptrons, is used to classify the data in this research. The original data sets are processed to have only the features that are selected by SVM-RFE or FCBF. The back propagation algorithm explained in Section 2.4.1 is used to train each of the ANNs. An example of an artificial neural network is shown in Figure 3.4 with five features as inputs, a single hidden layer and the output of 8 textile classes.

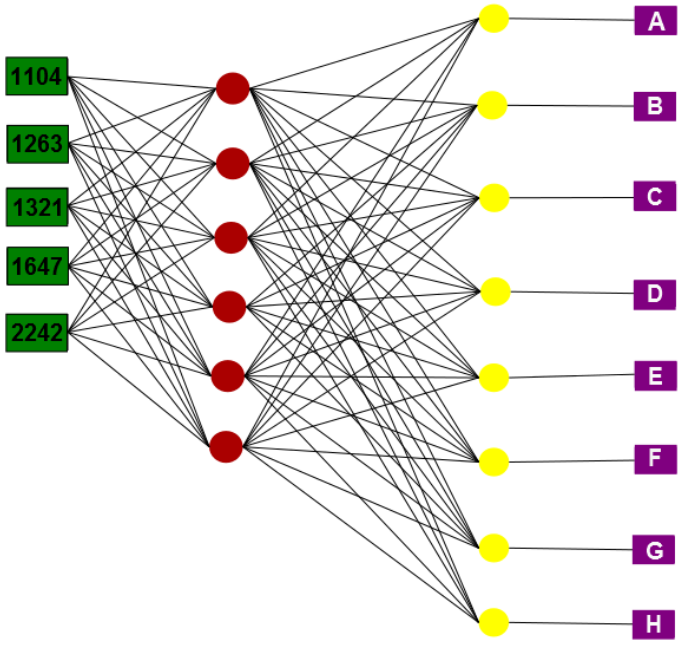


Figure 3.4: Example of an ANN (composition classification): green nodes are inputs with wavelength numbers, red nodes are in the hidden layer, yellow nodes are the output nodes, and purple labels on the right show the class the output node represents.

All nodes in the ANN use the sigmoid function as the activation function defined by [27]:

$$S(x) = \frac{1}{1 + exp^{-x}}, \tag{3.2}$$

where x is the input value from the data to the sigmoid function. A graph of a sigmoid function can be seen in Figure 3.5.

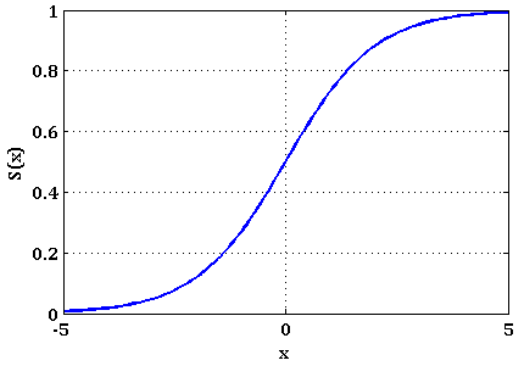


Figure 3.5: Sigmoid function where x is the input value, maps to the corresponding $S(x)$ of the activation function.

The number of nodes in each hidden layer n in the ANN is determined by [52]:

$$n = \frac{N + C}{2}, \tag{3.3}$$

where N is the number of features or inputs and C is the number of classes or outputs. The specific parameters for training each ANN are further described in Chapter 4.

Each ANN used in this thesis is trained using the K-fold cross validation method [15]. K-folds cross validation divided a training data set into K equal and disjoint sets. A classifier (ANN) is trained using the $K - 1$ disjoint subsets and tested on the remaining set. This process is repeated K times, each using a different $K - 1$ subsets for training. The default value of K, 10, is used for all training of classifiers. Epochs refer to the number of iterations for a given training cycle. The optimal epoch number for each classifier is

determined by performing an exhaustive number of iterations and selecting the epoch where the error levels off. This process is demonstrated in Section 4.1.

The classification performance of an ANN for a multi-class problem is calculated using an equal weighted accuracy (EWA) measure. EWA computes the average of the accuracies for each class by [16]:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\text{correctly classified instances of Class } i}{\text{size of Class } i}, \tag{3.4}$$

where i is the numeric class index and N is the total number of classes in the given multi-class problem. In the case of binary classification, the accuracy is calculated by [53]:

$$\frac{\text{True Positive + True Negative}}{\text{True Positive + False Positive + True Negative + False Negative}}. \tag{3.5}$$

An example binary classification contingency table is shown in Table 3.3, where the positions of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are depicted. FN rates and FP rates are also calculated as performance measures by [53]:

$$\text{FN rate} = \frac{\text{FN}}{\text{FN + TP}} \tag{3.6}$$

$$\text{FP rate} = \frac{\text{FP}}{\text{FP + TN}} \tag{3.7}$$

Table 3.3: Example contingency table of a binary classification showing locations of TP, FP, TN, and FN. A and B are arbitrary classes.

		Classified As	
		A	B
True Class	A	True Negative	False Positive
	B	False Negative	True Positive

3.2.4 Textile Fingerprinting Model.

The assumed textile fingerprinting model could be utilized as described in the following process. The model is built upon three different sets of features and ANNs. This process works with the assumption that the detectors already have a visual contact of the FOI and that the textile library has access to samples of the same textile type as that of the FOI. The initial screening of the FOI collects the features (wavelengths) necessary for color and composition classification. Two existing ANNs, each already trained for color and composition classification, processes the collected features, from the screening of the FOI to determine the FOI's color and composition. Depending on the color and composition classification of the FOI, the features needed for reacquisition of the FOI vary, consisting of one feature (wavelength) in the VIS spectrum for color discrimination and 2 to 4 features (wavelengths) in the NIR/SWIR for the uniqueness classification (next and last phase of fingerprinting) of a known textile. The secondary screening collects the information from the FOI for these features necessary for fingerprinting of the FOI. When the visual contact is compromised (*i.e.* the FOI goes into a building), a new ANN is trained using samples from the textile spectral library and the features collected from the secondary screening to reacquire the FOI. The newly trained ANN is then used to identify the FOI from the dismounts coming out of the building in the reacquisition stage. An example of the textile fingerprinting model is illustrated in Figure 3.6.

The textile fingerprinting model can be seen as a process of elimination, the last step being identification, using one versus all classification. A one versus all classification can be treated as a detection or binary classification problem, where the first class is the FOI and the second class is all other classes (textiles). The performance of a model is measured by the ability to identify a FOI. The model is trained with a random assortment of sample textiles to detect the FOI amongst a group of other textiles. The accuracy and performance of each stage of the textile fingerprinting model will be discussed in Chapter 4.

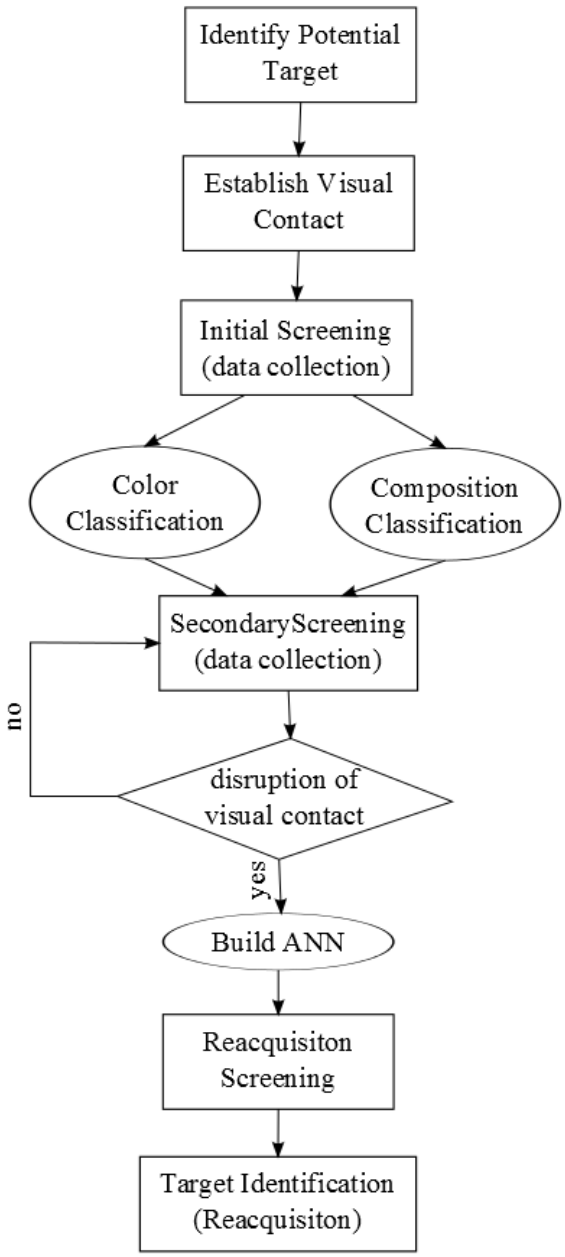


Figure 3.6: Flowchart of the textile fingerprinting process. First a potential target is identified and visual contact with target is established. Initial screening collects features for classification of color and composition. Secondary screening collects features according to the previous classification. An ANN is built to uniquely identify an FOI, when visual contact is disrupted, from features collected in the secondary screening. The ANN is used in the reacquisition screening for identification of the FOI.

3.3 Summary

Data collection, preprocessing, feature selection and classification make up the methodological backbone of this thesis. Spectral reflectance data of textile sample and color chips are collected using the FieldSpec ® 3 Hi-Res Spectroradiometer [2]. The visible and NIR/SWIR wavelengths are processed separately, the former for color and latter for textile composition. SVM-RFE is selected as the feature selection technique for color and composition classification. A computationally less expensive feature selection method, FCBF is chosen for textile uniqueness classification. A model for uniquely identifying (fingerprinting) an FOI is built using the features selected and artificial neural networks.

IV. Results

This chapter presents the results of the textile fingerprinting model outlined in Chapter 3. The three stages (color classification, composition classification, and textile uniqueness identification) of the fingerprinting process of textiles are discussed in detail. Three distinct reflectance data sets are collected: paint chips, pristine textiles, and altered textiles (washed and dried). Feature selection and classification (using ANNs) are implemented on all data sets for each stage of the fingerprinting model. The textile fingerprinting model, as illustrated in Figure 3.6, processes an FOI by classifying the color and composition of the FOI followed by collecting secondary data of the FOI. The reacquisition of the FOI consists of modeling a separate ANN for uniquely identifying the FOI. The feature selection and classification methods pertinent to each step are explained in the following sections.

The data collected, methodology implemented, and the results for color classification are discussed in Section 4.1. Section 4.2 discusses the pertinent information for composition classification and its results. The details regarding the textile uniqueness identification are presented in Section 4.3, and the overall performance of the fingerprinting model is reported in Section 4.4.

4.1 Color Classification

The hyperspectral reflectance data of each paint chip depicted in Figure 3.3 is collected with a ASD FieldSpec ® 3 Hi-Res spectroradiometer [2] using a contact probe. Each measurement is collected from 350nm - 2500nm with 1nm resolution providing 2151 features. Feature selection in the visible wavelengths for color classification is accomplished with 351 features from 350nm - 700nm.

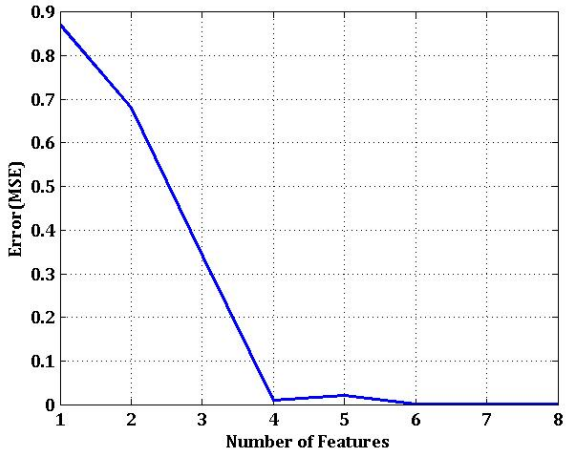


Figure 4.1: Number of features selected vs. mean squared error of classification of color with an ANN trained with the selected features. As number of features used increases, the MSE decreases.

There are eight distinct groups of colors in the 'color data set': blue, green, indigo, orange, purple, red, yellow, and turquoise. Each color group contains four different levels of shading (labeled by letters a, b, c, d) which gives 32 color classes. The shading is linearly divided between the four levels, where the 'a' class has the least amount of gray added and the 'd' class has the most amount of gray added. Features, for color classification, are selected from the entire data set using a Support Vector Machine based on the Recursive Feature Elimination (SVM-RFE) method in Weka.

The SVM-RFE algorithm ranks the features, best to worst, allowing the user to select the number of desired features. To determine the optimal number of features, an artificial neural network is trained and tested iterating from the greatest to least significant feature, increasing the number of features after each iteration. Figure 4.1 shows the relationship of the number of features versus the MSE; as the number of features increases, the MSE decreases. The top four features (530nm, 481nm, 588nm, and 430nm) are selected for

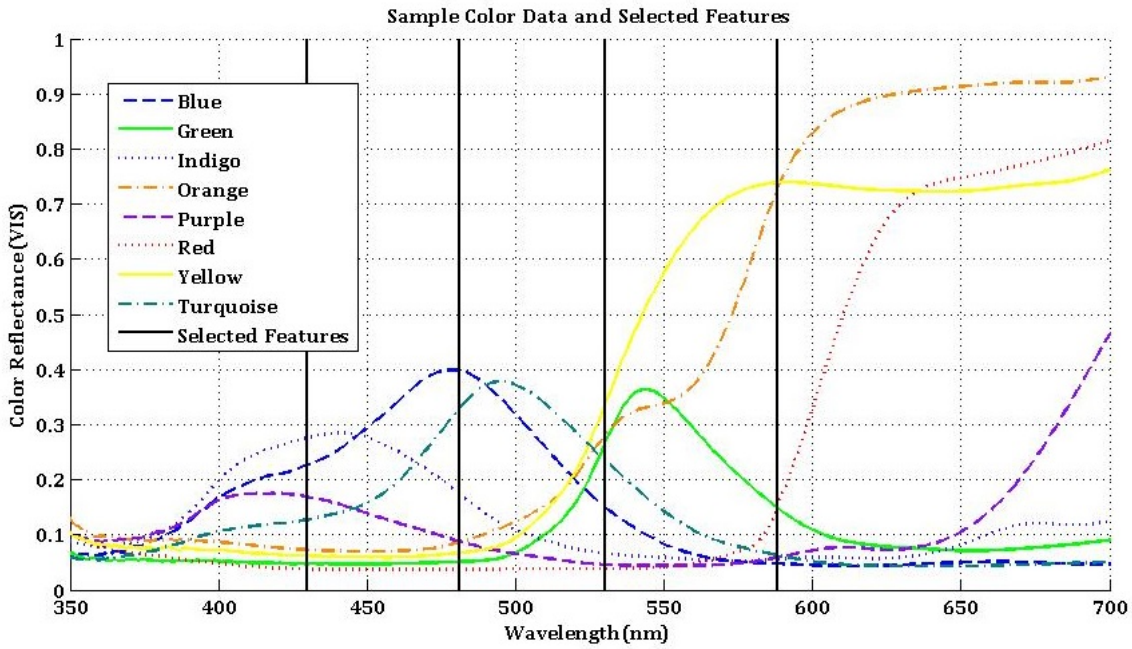


Figure 4.2: Selected features for color classification and the average reflectances for the eight a-class colors. The solid vertical black line represents the wavelength selected as a feature (430nm, 481nm, 530nm, 588nm).

color classification based on the results shown in Figure 4.1. The chosen wavelengths are depicted in solid black vertical lines, along with the average reflectances of the a-class colors in Figure 4.2.

Training of the artificial neural network is accomplished with the selected feature of the color data set. Relevant parameters used for the training of the ANN are reported in Table 4.1. The ANN is structured where the four features are the inputs, a single numeric output that allows distinction between 32 classes, and a single hidden layer is used to minimize the computation complexity of the back propagation algorithm. The number of neurons used in the hidden layer is calculated from Equation (3.3), and the Sigmoid, shown in Equation (3.2), is the activation function. The learning rate and momentum

Table 4.1: Parameters for the color classification ANN.

Parameter	**Value**
Inputs	4
Outputs	32
Hidden layers	1
Neurons in hidden layer	18
Activation function	Sigmoid
Learning rate	0.3
Momentum	0.2
Epochs	150
K (cross validation)	10

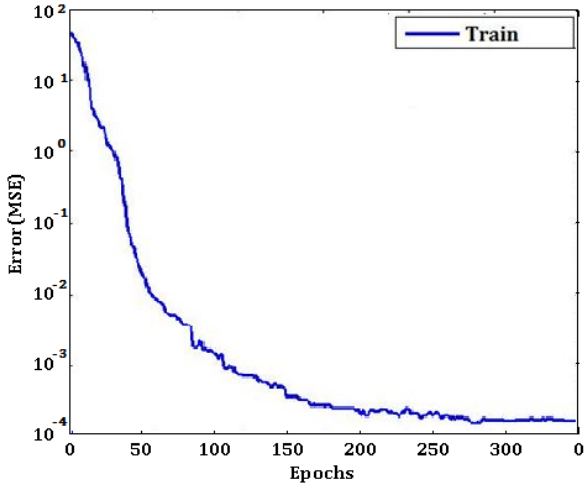


Figure 4.3: Illustrates the MSE for the first 300 epochs of the artificial neural network for color classification. The MSE gradually levels off after 150 - 200 epochs.

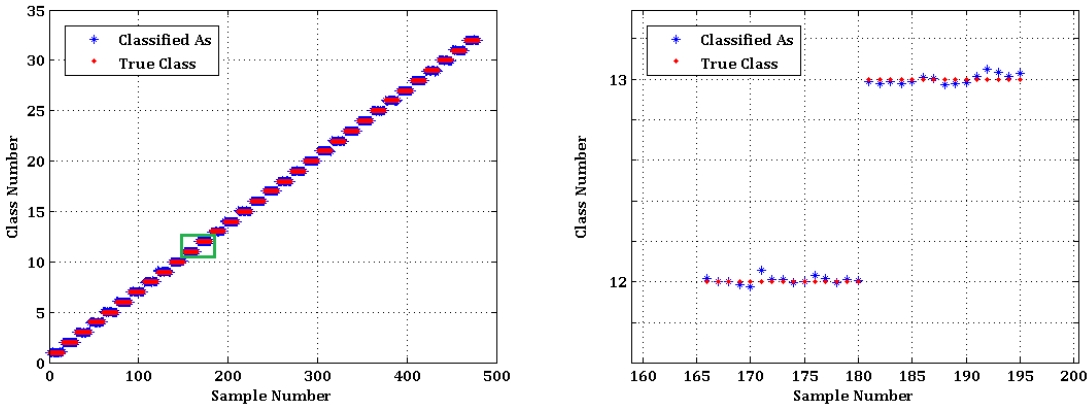
parameters, which dictate the amount of change to the weights during each iteration, are 0.3 and 0.2. An exhaustive number of training iterations are completed to determine an optimal number of epochs. Figure 4.3 indicates that 150 epochs is adequate for this particular training set. The 'K' parameter, of Table 4.1, refers to the number of K-folds cross validations performed.

The color classification performance is evaluated using a reserved validation data set that the artificial neural network is not permitted to use for training. Figure 4.4 shows the classification results of the validation data set. Each of the 32 color classes are designated by numeric class value of 1 through 32. The sample numbers associated with a numeric class are determined by:

$$(C-1)\times 15+1 \quad \sim \quad (C-1)\times 15+15, \tag{4.1}$$

where C is the numeric class number. For example, the sample numbers corresponding to class 12 would be samples 166 to 180. Each red dot, shown in Figure 4.4, is the true class of a sample and each blue star represents the class predicted by the model. As seen in Figure 4.4, all samples are classified into their corresponding numeric class, giving the color classification process 100% accuracy calculated by Equation (3.4). This result is expected due to the fact that these color classes are linearly separable.

The features associated with the 32 class distinction process are necessary for initial determination of the FOI's color. Reacquisition of the FOI is then accomplished based on the base color (8 class) which requires only one feature for discrimination. A single feature is selected to discriminate between the classes (a,b,c,d) of each group using SVM-RFE. The chosen feature for each color group is recorded in Table 4.2, which is later used in the final textile identification process.



(a) (b)

Figure 4.4: The left graph (a) shows the classification results of the validation color data set. The y-axis represents the number coded color class and the x-axis represents the sample number. Graph (b) shows the zoomed in version of the green box of graph (a) for the classification for class number 12 and 13.

Table 4.2: Selected feature for within-class color classification

Color	**Wavelength(nm)**	**Color**	**Wavelength(nm)**
Red	636	Blue	492
Orange	587	Indigo	453
Yellow	563	Turquoise	500
Green	536	Purple	422

4.2 Composition Classification

The textile samples are collected from 750nm - 2500nm, with 1nm resolution, for composition feature selection and classification. The pristine textile data set is processed with the SVM-RFE method, where the optimal feature set is determined by training and

testing an ANN with an increasing number of features, starting with the most significant feature. Figure 4.5 depicts the effect of increasing the number of features on the MSE of the composition classification.

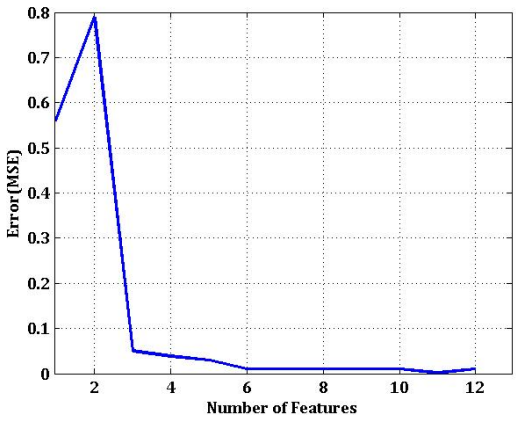


Figure 4.5: Number of features selected vs. mean squared error of composition classification by an ANN trained with selected features. As the number of features increases, the MSE decreases. The spike at 2 features can be explained by the presence of noisy samples or outliers since this is not an average over multiple trials.

The top five ranked features reported from the SVM-RFE method are selected for composition classification due to the results shown in Figure 4.5. A global feature set of the composition classification for the textiles used in this thesis consist of five features: 1104nm, 1263nm, 1324nm, 1647nm, 2242nm. The selected features (solid black vertical lines) are shown in Figure 4.6 along with the average reflectance measurements of each pristine textile sample.

An ANN is trained for the classification of textile composition using the parameters listed in Table 4.3. The activation function, learning rate, and momentum parameters used are the same as those in the color classification model. The composition classification ANN accepts five features as inputs, contains seven neurons in a single hidden layer, and has a numeric output that allows distinction between eight output classes. The

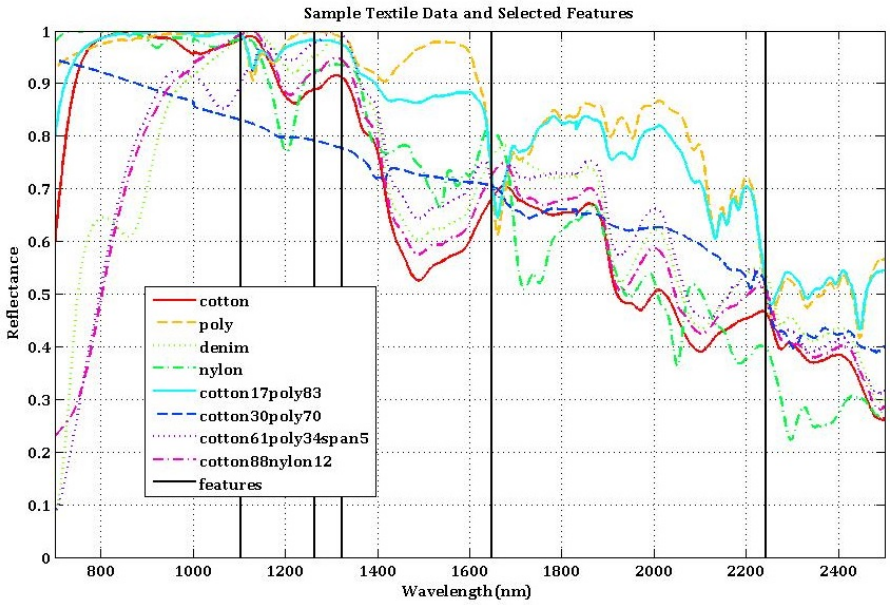


Figure 4.6: Selected features for composition classification and the average reflectance measurements of the eight textiles. The solid black vertical lines represent the features (wavelengths) selected: 1104nm, 1263nm, 1324nm, 1647nm, 2242nm.

classification performance, evaluated by testing the trained ANN with a validation data set, is reported in the confusion matrix shown in Table 4.4. The 100% accuracy classification result, calculated with Equation (3.4), for the textile composition classification is not surprising since this is easily explained by the small data set of only eight textile classes, which are separated by large margins in the features selected by SVM-RFE. It can be seen in Figure 4.6 that two of the selected features (1256nm and 1324nm) are able to linearly separate a bulk of the textile classes.

Table 4.3: Parameters for the composition classification ANN.

Parameter	**Value**	**Parameter**	**Value**
Inputs	5	Learning rate	0.3
Outputs	8	Momentum	0.2
Hidden layers	1	Epochs	500
Neurons in hidden layer	7	K (cross validation)	10
Activation function	Sigmoid		

Table 4.4: Confusion Matrices for Composition Classification. The left matrix shows classification results from the training and the right matrix shows results from the validation data set. Both classifications result in a 100% accuracy. Each letter corresponds to a textile composition type (A: 100% cotton, B: 100% polyester, C: 100% cotton-denim, D: 100% nylon, E: 17% cotton 83% poly, F: 30% cotton 70% poly, G: 61% cotton 34% poly 5% spandex, H: 88% cotton 12% nylon). The diagonals correspond to the accurate classifications. As seen in both the training and testing results, there are no misses or false alarms for the classification of textile composition.

Training

True Class \ **Classified As**	A	B	C	D	E	F	G	H
A	50	0	0	0	0	0	0	0
B	0	50	0	0	0	0	0	0
C	0	0	50	0	0	0	0	0
D	0	0	0	50	0	0	0	0
E	0	0	0	0	50	0	0	0
F	0	0	0	0	0	25	0	0
G	0	0	0	0	0	0	50	0
H	0	0	0	0	0	0	0	50

Testing (Validation)

True Class \ **Classified As**	A	B	C	D	E	F	G	H
A	50	0	0	0	0	0	0	0
B	0	50	0	0	0	0	0	0
C	0	0	50	0	0	0	0	0
D	0	0	0	50	0	0	0	0
E	0	0	0	0	50	0	0	0
F	0	0	0	0	0	25	0	0
G	0	0	0	0	0	0	50	0
H	0	0	0	0	0	0	0	50

4.2.1 Feature Selection Comparison.

To further evaluate the performance of the five feature global set, the SVM-RFE feature selection method is compared with other existing methods: ReliefF [42], FilterSubsetEval with best first search (BFS) [52], and WrapperSubsetEval with BFS [52]. The top five features are selected from each feature selection method, unless a smaller subset of features is returned by the algorithm. The selected features are displayed in Table 4.5 as well as the classification accuracy, based on an ANN that is trained and tested with the corresponding feature set.

Table 4.5: The features, accuracies, and runtime of the different feature selection methods. Four feature selection methods are tested by training and testing an ANN with the selected feature set. The accuracies reported are calculated with samples correctly classified as that class. Runtime is specific to a system with an Intel Core i7-3610M 2.4 GHz processor.

Method	**Features (nm)**	**Accuracy (%)**	**Runtime (sec)**
SVM-RFE	1104, 1263, 1324, 1647, 2242	100	13048
ReliefF	828, 827, 829, 826, 825	86	123
FilterSubsetEval	1421	52	13
WrapperSubsetEval	1235	57	392

As discussed in Section 2.3.1, ReliefF generates non-optimal feature sets with redundant or highly correlated features [42]. The ANN trained with features selected with ReliefF returns an 86% accuracy. This result may be explained by the fact that the ReliefF algorithm returns a group of consecutive features, as the algorithm uses nearest neighbor comparisons when selecting features and does not remove redundant features.

FilterSubsetEval evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy between

them [52]. WrapperSubsetEval evaluates feature sets by using a user defined learning scheme, BFS [52]. Cross validation is used to estimate the accuracy of the learning scheme for a set of attributes. Both filter and wrapper methods use BFS to find subsets which searches the space of features by a greedy hill-climbing method. However, hill-climbing algorithms are often subject to local minima or maxima [15]. The filter and wrapper methods combined with BFS return a single wavelength as the optimal subset of features, which may be due to the hill-climbing nature of the search algorithm. An ANN is trained with the selected feature of each method, and the performance of each model is verified with a validation data set, calculating the equal weighted accuracy of the outputs. The accuracies are reported in Table 4.5 as 52% and 57%. As expected, classification systems with a single feature result in unsatisfactory accuracies, performing only slightly better than random guessing.

The runtime of each algorithm is also used to compare the performance of each feature selection method. Runtime is used as a relative measure of performance as the runtime will be different for every system used. The reported runtimes are specific to a system geared with an Intel Core i7-3610M 2.4 GHz processor. Of the four feature selection methods, the SVM-RFE has the longest runtime, exceeding 3 hours, whereas the other methods are completed under 10 minutes. The extended runtime of SVM-RFE results from its recursive nature as it trains the SVM repeatedly. However, since the feature selection for textile composition is completed off-line and only needs to be run once, the runtime of an algorithm can be neglected, making the accuracy performance more important than the computation simplicity. Disregarding the runtime, SVM-RFE outperforms all three feature selection algorithms, which could be caused by the complexity of the more sophisticated selection algorithm. Also, the performance of the four feature sets were tested only with ANNs, which may not have given an unbiased evaluation of each selection algorithm.

4.3 Textile Uniqueness Classification

Each sample of the same make (*i.e.* 100% cotton) is divided into four sections and treated with various combinations of machine washing, machine drying, and air drying. Four classes are created for each textile group and are labeled as: pristine, washed and air dried 5 times, washed and machine dried 5 times, and washed and machined dried 20 times.

The eight textile groups are processed separately by the fast correlation-based filter (FCBF) in Weka. FCBF is used for feature selection of textile uniqueness classification due to the fact that the algorithm is less complex and faster to implement. Processing every single textile type with SVM-RFE is computationally expensive and timely. The FCBF feature selection algorithm returns a small subset of optimal features, unlike the SVM-RFE; therefore, the resulting subset is used as the optimal feature set. For example, Figure 4.7 illustrates the location of the selected features relative to the textile reflection signature of 100% cotton. The selected features for the other seven textile groups can be found in Appendix A. The wavelengths selected for each textile group are presented together in Table 4.6.

The subset of optimal features found by FCBF varies from 2 to 4 features, depending on the textile make. The different number of optimal subsets could be explained by the fact that the same threshold was used to determine the relevance of each feature for all textile types. The number of features in the optimal set is different for each textile make because the FCBF algorithm may deem more features relevant for one textile make than another.

Table 4.7 shows the parameters used when training the artificial neural networks for textile identification. Inputs of each ANN vary by the number of features used, f, for each textile group, as well as the neurons in the hidden layer, based on Equation (3.3). The learning rate, momentum and number of K cross validations are kept at the default values set by Weka: 0.3, 0.2, and 10. A sigmoid is used as the activation function, and the

Table 4.6: Features selected for each textile group by FCBF, using four unique versions of each textile.

Textile Type	**Features(nm)**
100% cotton	1271 1830 2292 1611
100% polyester	1333 2346
100% cotton-denim	890 1427 1829 2020
100% nylon	1213 1727
17% cotton 83% poly	800 1353 1907 2345
30% cotton 70% poly	1093 1666
61% cotton 34% poly 5% span	816 1426 1875 2004
88% cotton 12% nylon	1052 1404 1940

number of epochs is determined as previously discussed in Section 3.2.3. An ANN is trained with half of the data set and validated with the other half. The performance of the classification of the validation data set can be seen for each textile group in Table 4.8. For the simplicity of the confusion matrices, each textile group's classes are designated by letter, A, B, C, and D, where group A is the pristine textiles, B is the textiles washed and air dried 5 times, C is the textiles washed and machine dried 5 times, and D is the textiles washed and machine dried 20 times. The overall validation equal weighted accuracy, calculated as described in Equation (3.4), of textile uniqueness classification is reported in Table 4.8, which averages to 95%.

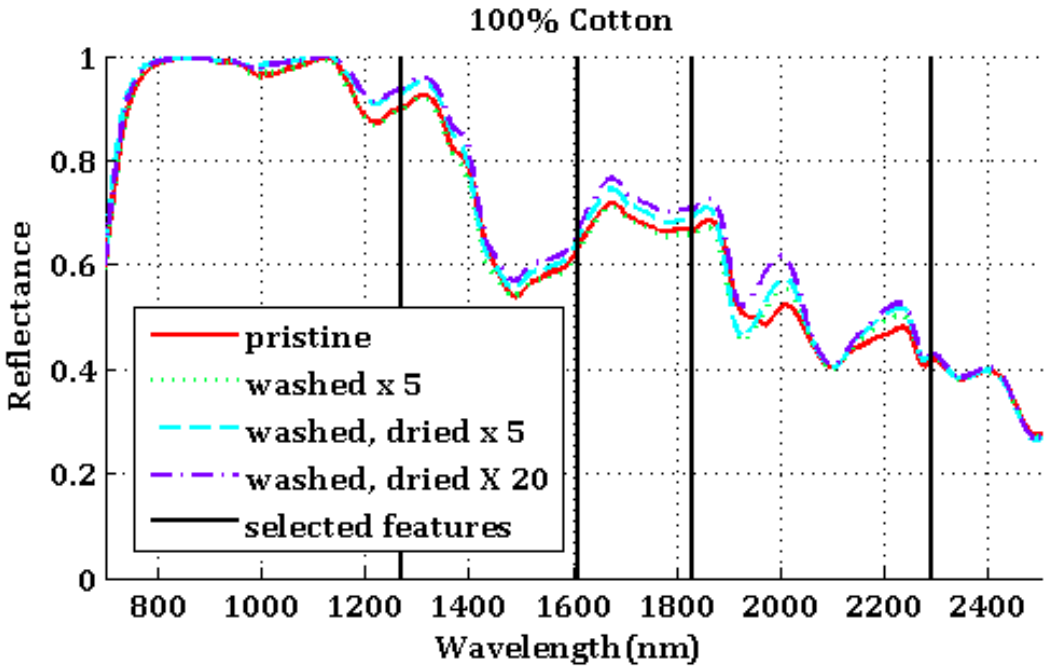


Figure 4.7: Selected features shown with black solid vertical lines, for the identification of 100% cotton and the average of each version of the 100% cotton textile (1271nm, 1830nm, 2292nm, 1611nm).

Table 4.7: Parameters for the textile uniqueness identification ANN.

Parameter	**Value**
Inputs	f
Outputs	4
Hidden layers	1
Neurons in hidden layer	$\frac{f+4}{2}$
Activation function	Sigmoid
Learning rate	0.3
Momentum	0.2
Epochs	500
K (cross validation)	10

Table 4.8: Confusion matrices for each textile uniqueness classification for the validation data set, classified by an ANN. The rows correspond to the true class and columns indicate classification results. Each version of textiles is labeled by A-pristine, B-washed, air dried × 5, C-washed, machine dried × 5, D-washed, machine dried × 20. Accuracies are calculated with EWA and are boxed.

Type/Accuracy

100% cotton

100%

True Class \ Classified As	A	B	C	D
A	50	0	0	0
B	0	50	0	0
C	0	0	50	0
D	0	0	0	50

Type/Accuracy

100% poly

90%

True Class \ Classified As	A	B	C	D
A	50	0	0	0
B	0	50	0	0
C	0	0	50	0
D	18	0	0	32

100% cotton

denim

89.5%

True Class \ Classified As	A	B	C	D
A	50	0	0	0
B	0	29	0	21
C	0	0	50	0
D	0	0	0	50

100% nylon

94.5%

True Class \ Classified As	A	B	C	D
A	50	0	0	0
B	0	50	0	0
C	0	11	39	0
D	0	0	0	50

17% cotton

83% poly

94.5%

True Class \ Classified As	A	B	C	D
A	50	0	0	0
B	0	50	0	0
C	11	0	39	0
D	0	0	0	50

30% cotton

70% poly

100%

True Class \ Classified As	A	B	C	D
A	25	0	0	0
B	0	50	0	0
C	0	0	50	0
D	0	0	0	25

61% poly

34% cotton

5% spandex

98%

True Class \ Classified As	A	B	C	D
A	50	0	0	0
B	0	46	0	4
C	0	0	50	0
D	0	0	0	50

88% cotton

12% nylon

99%

True Class \ Classified As	A	B	C	D
A	48	0	0	2
B	0	50	0	0
C	0	0	50	0
D	0	0	0	50

4.4 Textile Fingerprinting Model Performance

To test the overall performance of the model built for textile fingerprinting, three fabrics of interest (FOIs) are selected from the existing library of textiles (five times washed and machine dried versions of 100% cotton, 100% polyester, and denim). Measurements of three new FOIs, never incorporated by any feature selection or ANN training process, are added to the existing textile spectral library to act as possible confusers. These additional textile samples are depicted in Figure 4.8. The respective reflectance measurements of the additional textiles can be found in Appendix B.

(a) (b) (c)

Figure 4.8: The first picture (a) shows a green 100% cotton shirt, washed, machine dried, and worn for 2 years. The middle (b) shows a purple 100% polyester shirt, washed, machine dried, and worn for 1 year. The last picture (c) shows a denim, 100% cotton jean, washed, machine washed, and worn for 3 years.

The four visible wavelengths and five NIR/SWIR wavelengths of all three FOIs are collected for classification by ANNs trained specifically for color and composition discrimination. When evaluated by the color and composition classification steps, the FOIs are classified as shown in Table 4.9. All color and composition classifications achieved a 100% accuracy.

Table 4.9: The true class and color and composition classification of the three FOIs. Each FOI is classified by the color discrimination ANN and composition discrimination ANN.

	Truth		**Classified As**	
FOI	**Color**	**Composition**	**Color**	**Composition**
1	green	100% cotton	green	100% cotton
2	red	100% polyester	red	100% polyester
3	blue	100%cotton - denim	blue	100%cotton - denim

Two distinct data sets are created for the purpose of training and testing of the one versus all textile fingerprinting ANN for each FOI. The sample allocations are shown in Table 4.10. The 50 FOI samples for training and testing are separated into two groups randomly, therefore making the two data sets entirely different, without overlap. The 400 non-FOI samples for training consist of two textiles of the same composition of the FOI, and two randomly selected textiles of a different type than the FOI. For example, FOI 1 is trained with the FOI 1 samples, the pristine 100% cotton, the 100% cotton washed and dried 20 times, the 88% cotton 12% nylon textile, and the 100% polyester textile. The 1050 non-FOI samples for training consist of all 32 textiles from the spectral library.

After an FOI is classified to a color and composition, it is uniquely identified. The wavelengths collected for the identification process are different for each FOI and are shown in Table 4.2 and Table 4.6. For example, if an FOI is classified as green and 100% cotton, wavelength 536nm for color, and wavelengths 1271nm, 1830nm, 2292nm, and 1611nm for composition are selected for textile uniqueness identification. To test the performance of the identification process, an ANN is trained to find one FOI from a group of five textile subjects as described in Table 4.10. The ANN is trained using 10-folds cross validation (default setting in Weka) in a one versus all approach. This model is then tested

Table 4.10: Description of training and testing (validation) sets for the textile fingerprinting model. The non-FOI samples for training include two randomly selected textiles of the same composition of the FOI, and two randomly selected textiles from the textile spectral library of a different type of the FOI. The non-FOI samples for testing contains 32 textiles of the existing spectral library. The 50 FOI samples used in training and testing are two different sets, randomly divided from a 100 sample pool.

Data Set	**Sample Allocation**
Training	50 FOI
	400 non-FOI (4 textile classes)
Testing	50 FOI
	1050 non-FOI (32 textile classes)

to find the same FOI from a larger group of 32 subjects. The training and testing classification results for each FOI are reported in Table 4.11.

The training and testing classification performance statistics are shown in Table 4.12. FOI 1, 2, and 3 each result in 99.4%, 97.2%, and 98.3% accuracy respectively when tested with the 32 class testing group. These accuracy numbers are calculated by methods described in Equation (3.5). However, these seemingly stellar accuracy results are inflated by the small size of the FOI sample compared to the testing set of 1100 samples. The testing accuracies are higher than the training accuracies, which is unexpected. This could be an effect from dividing the FOI samples into two separate 50 sample sets each for training and testing. Since the FOI samples for training and testing are entirely different, a higher testing accuracy is feasible.

There are 20 false alarms and 10 misses in the case of FOI 2, even with a 97.2% classification accuracy. False negative (FN) and false positive (FP) rates calculated by

Table 4.11: Training and Testing confusion matrices for each FOI, trained with 450 samples and tested with 1100 samples.

Training

True Class \ Classified As	FOI1	non-FOI1
FOI1	395	5
non-FOI1	3	47

True Class \ Classified As	FOI2	non-FOI2
FOI2	381	19
non-FOI2	11	39

True Class \ Classified As	FOI3	non-FOI3
FOI3	386	14
non-FOI3	14	36

Testing (Validation)

True Class \ Classified As	FOI1	non-FOI1
FOI1	1046	4
non-FOI1	2	48

True Class \ Classified As	FOI2	non-FOI2
FOI2	1030	20
non-FOI2	10	40

True Class \ Classified As	FOI3	non-FOI3
FOI3	1037	13
non-FOI3	5	45

Equation (3.6) and Equation (3.7) show a more in-depth performance review of the one versus all ANN for FOI fingerprinting. For instance, the training FN rate of FOI 2 and FOI 3 are at 0.28 and 0.38, which indicates that the FOI is missed at a higher rate than one out of four.

Table 4.12: Training and testing classification performance statistics including accuracy, false negative (FN) rate, and false positive (FP) rate for each FOI. Each statistic is calculated by methods explained in Section 3.2.3.

	Training			**Testing**		
	Accuracy (%)	FN rate	FP rate	Accuracy (%)	FN rate	FP rate
FOI 1	98.2	0.06	0.01	99.4	0.04	0.003
FOI 2	93.3	0.28	0.05	97.2	0.20	0.02
FOI 3	93.7	0.38	0.04	98.3	0.11	0.01

4.5 Summary

This chapter reviewed the results of two feature selection methods, SVM-RFE and FCBF, to distinguish between color, composition and finally uniquely classify (fingerprint) textiles. A neural network at each stage of the identification process is trained, using the features selected by the two methods. The textile fingerprinting model is able to classify 32 different colors, 8 types and blends of textiles, and identify between 4 versions of each textile. The overall performance of the fingerprinting model is tested by three FOIs from the existing textile spectral library with additional samples never seen by the model to act as confusers.

V. Conclusion and Future Work

The technology for identifying a dismount has numerous applications such as search and rescue, surveillance, and security purposes. Traditional dismount detection and identification efforts have included skin detection and thermal imaging [3, 57]. Although clothing is not an innate biological feature of a dismount such as skin, iris, fingerprint, and hair, it can provide more information about a dismount. Accurate textile identification can aid in developing a more robust dismount detection system. The ability to uniquely identify textiles relies on distinguishing between different material types and blends. Hyperspectral data provides characteristics of materials from the visible (VIS) to short-wave infrared (SWIR) wavelengths, which allows for detection of materials using their spectral signatures. When using hyperspectral data for material detection, it is crucial to identify the information that is effective in discriminating between materials. Feature selection methods are designed to select features of highly discriminating capability without redundancy or correlation.

The goal of this thesis is to determine a feature selection and classification process that is ideal for uniquely identifying textiles using their hyperspectral fingerprint. An optimal feature selection method that returned a global set that could uniquely identify all types and versions of textiles was not created or discovered; instead, a three step process of textile fingerprinting is presented.

In this chapter, the summary and conclusions of the work accomplished on uniquely identifying textiles are presented. Recommendations for future work related to fingerprinting textiles at the VIS - SWIR wavelengths and improvements that can be made on the current model are discussed.

5.1 Summary of Results

This thesis presented a hierarchical model to uniquely identify textiles using material reflectance information from the VIS, NIR, and SWIR regions. Of the feature selection methods reviewed and tested, Support Vector Machines with the aid of Recursive Feature Elimination (SVM-RFE) and fast correlation-based filter (FCBF) are utilized. Artificial neural networks (ANN) with multi-layer perceptrons (MLP) are trained and tested based on the selected features to classify the textiles. The SVM-RFE is shown to outperform other feature selection algorithms such as ReliefF, FilterSubsetEval, and WrapperSubsetEval with best first search. The MLPs consistently produced better accuracy classification results as compared to Naive-Bayes, SVMs, and radial basis function (RBF) networks.

The model developed can be divided into three major parts: color classification, composition classification, and textile uniqueness classification. Here, composition refers to the material types and blends of a specific textile sample, whereas uniqueness refers to the specific version of a textile make. The classification for both color and composition with validation data sets produce 100% accurate results. The validation results for textile uniqueness classification within each textile group averages out to give an overall 95% accuracy as seen in Section 4.3. Over training, or over fitting, is always a concern with learning algorithms; however, K-folds cross validation is used and textile samples are measured from different regions of the shirts in efforts to minimize such occurrences.

Depending on the color and composition of a fabric of interest (FOI), a different set of wavelengths are selected for the textile identification phase. For example, if an FOI is classified to be yellow and 100% polyester, wavelengths 563nm, 1333nm, and 2346nm are selected to train an ANN for identifying the FOI. The first wavelength, 563nm, discriminates between the different shades of yellow, and the two latter wavelengths, 1333nm and 2346nm, discriminate between the different versions of 100% polyester.

Once the color and composition of a sample are known, the detector is able to identify that specific sample from a pool of diverse textiles. Tests with three selected FOIs from the existing spectral library show that this textile fingerprinting model identifies textiles with 99.4, 97.2, and 98.3 percent accuracy within the eight textile class library created in this thesis.

5.2 Recommendations for Future Work

Future work with the model for uniquely identifying textiles proposed would involve a field data collect with the use of an actual hyperspectral imager (HSI). All reflectance measurements collected for training and validating the classifiers were from the ASD spectrometer with a contact probe that eliminated noise and atmospheric effects. Collecting HSI data with textiles will allow for the determination of the capability of this model in the field as well as for identifications of possible confusers that would cause false alarms in the system. To use the field data to train and test the model, the data will have to be treated for atmospheric correction and noise suppression.

Another direction the work in this thesis could be further explored is incorporating more types of fabrics and more versions within a textile group. The current model is only able to discriminate between eight different types and blends of textiles. This data set of eight, however, is not representative of the wide variety of possible types and blends of fabrics in the real world. The inclusion of more types of textiles will serve two distinct purposes. It will test the robustness of the model already designed, as well as aid in the determination of adding or eliminating wavelengths from the feature set to better classification accuracy.

The textile identification phase of the current model is tested at most with four different versions of the same textile group. To make separations between the versions of textiles, the samples were washed and dried a different number of times. Introducing different types of washing and drying machines as well as detergents and fabric softeners to textile

samples may shed light on new features with more discriminating capability. Further analysis of the results of the textile uniqueness classification with appropriate station keeping of samples will also shed light on why the false alarms and misses are made by the ANNs. Exploration of other feature selection algorithms as well as classification methods are also recommended for textile uniqueness classification as only FCBF and ANN were implemented in this thesis.

The question of a single global feature set still remains unanswered. If possible, a true global feature set would be able to identify an FOI of any class with high accuracy without jumping through hoops of classifying its color or composition first. A diverse training data set with more types of fabrics and separation would be a good start to a search for a global feature set.

Appendix A: Selected Features for Uniqueness Classification of Textiles

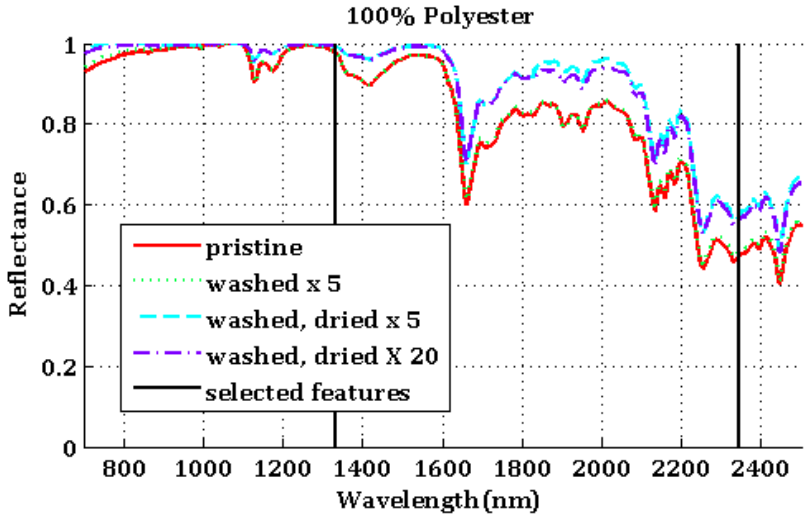


Figure A.1: Selected features shown with black solid vertical lines, for the identification of 100% polyester and the average of each version of the textile (1333nm, 2346nm).

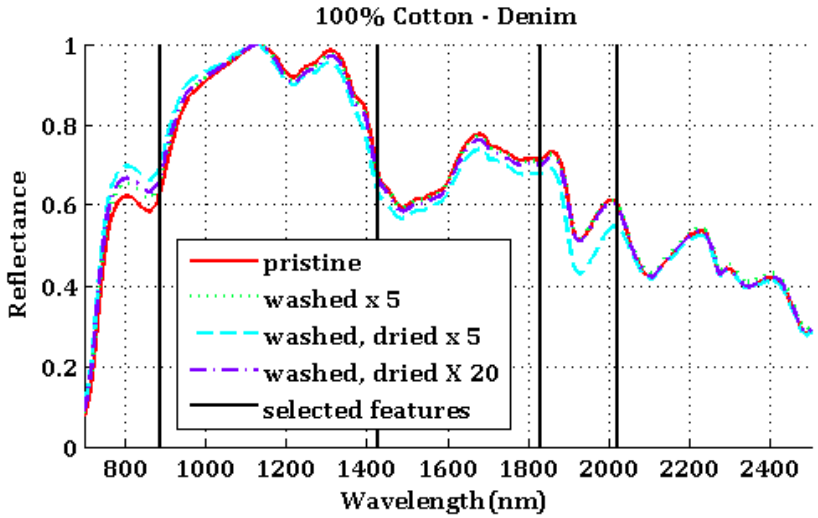


Figure A.2: Selected features shown with black solid vertical lines, for the identification of denim, 100% cotton and the average of each version of the textile (890nm, 1427nm, 1827nm, 2020nm).

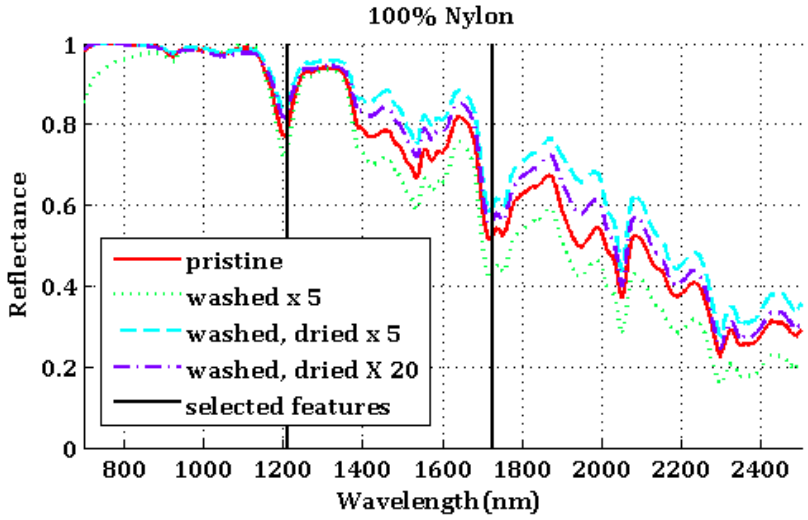


Figure A.3: Selected features shown with black solid vertical lines, for the identification of 100% nylon and the average of each version of the textile (1213nm, 1727nm).

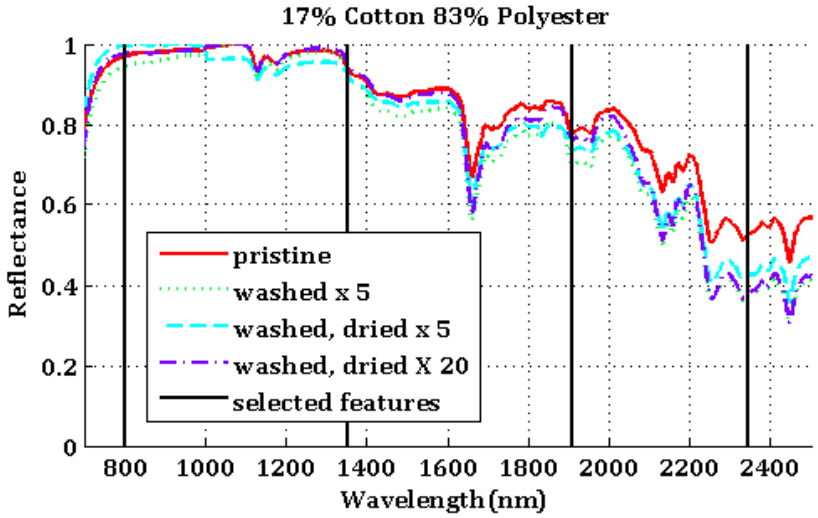


Figure A.4: Selected features shown with black solid vertical lines, for the identification of 17% cotton 83% polyester blend and the average of each version of the textile (800nm, 1353nm, 1907nm, 2345nm).

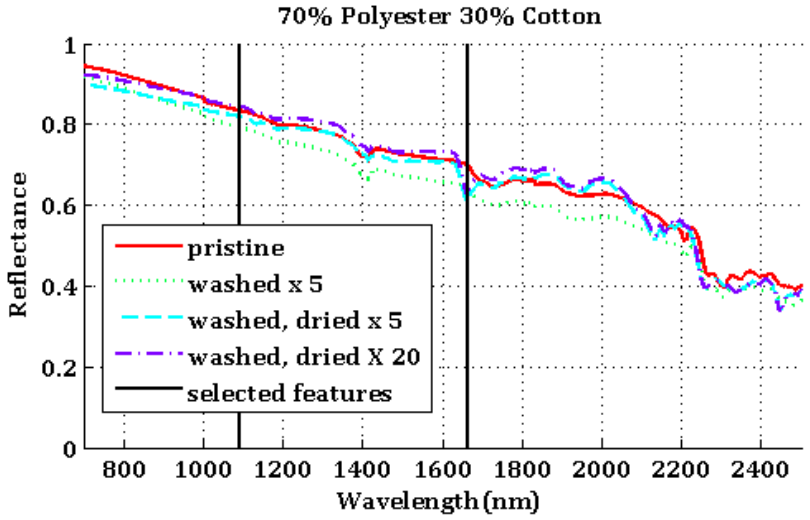


Figure A.5: Selected features shown with black solid vertical lines, for the identification of 30% cotton 70% polyester blend and the average of each version of the textile (1093nm, 1666nm).

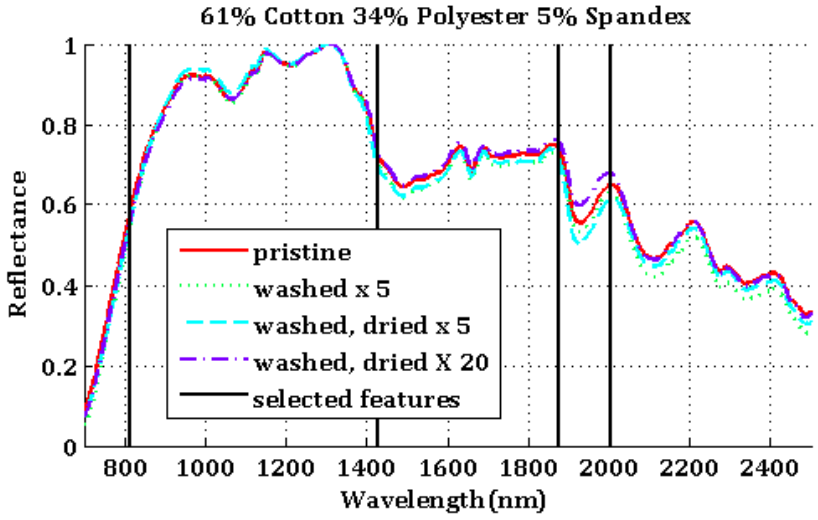


Figure A.6: Selected features shown with black solid vertical lines, for the identification of 61% cotton 34% polyester 5% spandex blend and the average of each version of the textile (816nm, 1426nm, 1875nm, 2004nm).

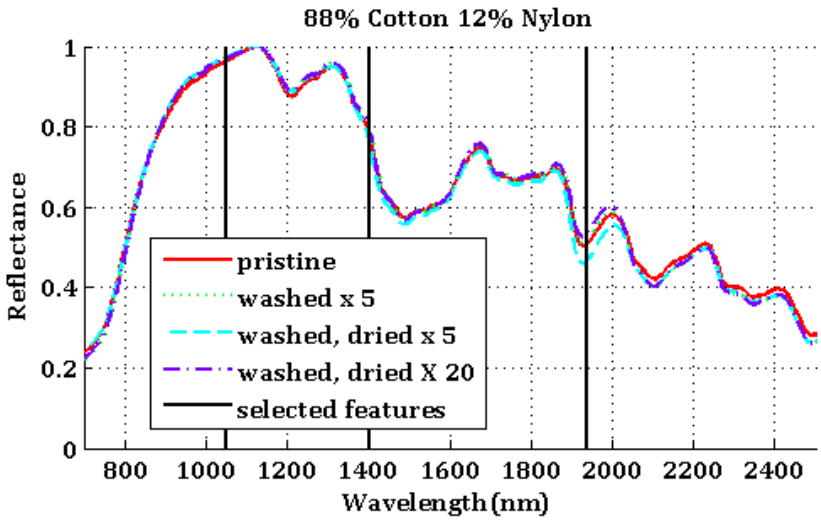


Figure A.7: Selected features shown with black solid vertical lines, for the identification of 88% cotton 12% nylon blend and the average of each version of the textile (1052nm, 1404nm, 1940nm).

Appendix B: Additional Fabric of Interest (FOI)

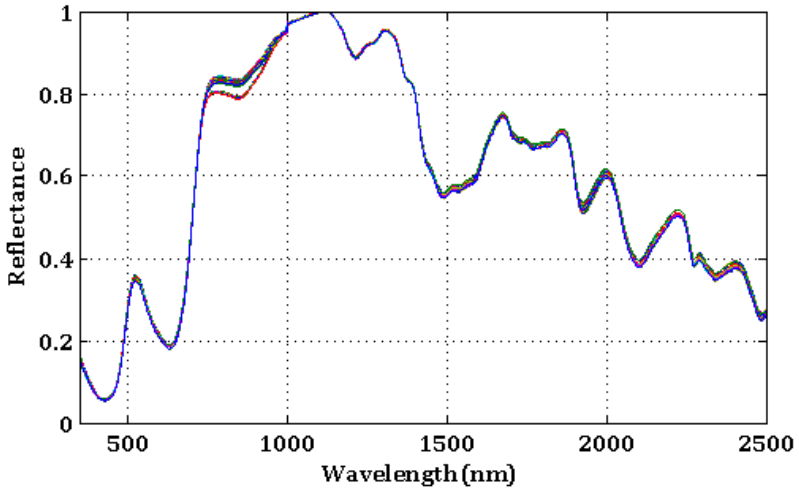


Figure B.1: 100 samples of full reflectance measurement of a green 100% cotton shirt, washed, machine dried, and worn for 2 years

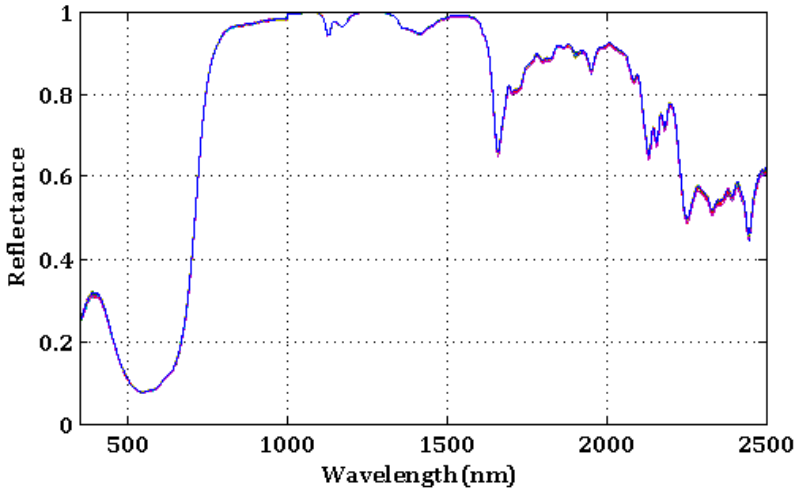


Figure B.2: 100 samples of full reflectance measurement of a purple 100% polyester shirt, washed, machine dried, and worn for 1 year.

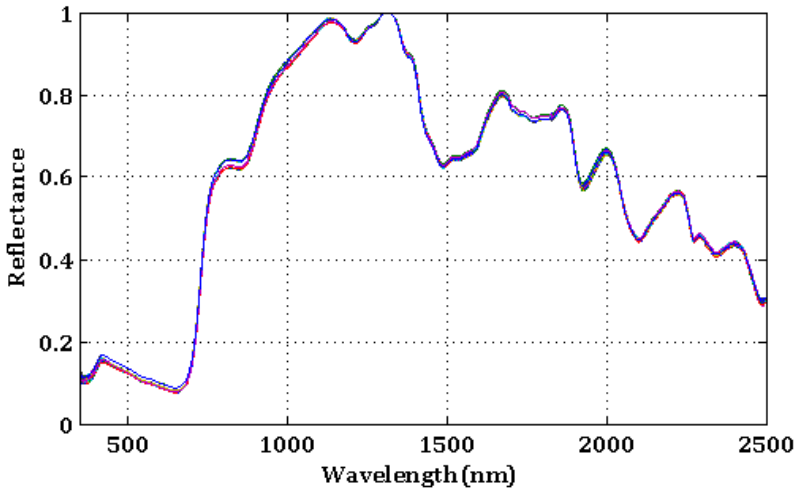


Figure B.3: 100 samples of full reflectance measurement of 100% cotton jean, washed, machine dried, and worn for 3 years.

Bibliography

[1] Agarwal, Abhishek, Tarek A El-Ghazawi, Hesham M El-Askary, and Jacqueline J Le-Moigne. "Efficient Hierarchical-PCA Dimension Reduction for Hyperspectral Imagery". *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, 353–356. 2007.

[2] ASD Inc. *FieldSpec 3 User Manual*, 1st edition.

[3] Beisley, Andrew P. *Spectral Detection of Human Skin in VIS-SWIR Hyperspectral Imagery without Radiometric Calibration*. Technical report, DTIC Document, 2012.

[4] Blackburn, Joshua, Michael Mendenhall, Andrew Rice, Paul Shelnutt, Neil Soliman, and Juan Vasquez. "Feature aided tracking with hyperspectral imagery". *Signal and Data Processing of Small Targets 2007*, 66990S–66990S, 2007.

[5] Blagg, A. "People detection using fused hyperspectral and thermal imagery". *SPIE Security+ Defence*, 854221–854221. International Society for Optics and Photonics, 2012.

[6] Blough, "Canal street crowded sidewalk Chinatown NYC", John.

[7] Borengasser, Marcus, William S Hungate, and Russell Watkins. *Hyperspectral remote sensing: principles and applications*. Crc Press, 2010.

[8] Brewer, Luke N, James A Ohlhausen, Paul G Kotula, and Joseph R Michael. "Forensic analysis of bioagents by X-ray and TOF-SIMS hyperspectral imaging". *Forensic science international*, 179(2):98–106, 2008.

[9] Briggs, David. "The Dimensions of Color", 2012. URL http://www.huevaluechroma.com.

[10] Brooks, Adam L. *Improved multispectral skin detection and its application to search space reduction for dismount detection based on histograms of oriented gradients*. Technical report, DTIC Document, 2010.

[11] Bruce, Lori Mann and Jiang Li. "Wavelets for computationally efficient hyperspectral derivative analysis". *Geoscience and Remote Sensing, IEEE Transactions on*, 39(7):1540–1546, 2001.

[12] Caputo, Barbara, Eric Hayman, and P Mallikarjuna. "Class-specific material categorisation". *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, 1597–1604. IEEE, 2005.

[13] Cheriyadat, Anil M. and Lori M. Bruce. "Why principal component analysis is not an appropriate feature extraction method for hyperspectral data". *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, volume 6, 3420–3422 vol.6. July.

[14] Clark, Jeffrey D. *Distributed Spacing Stochastic Feature Selection and its Application to Textile Classification*. Technical report, DTIC Document, 2011.

[15] Clark, Jeffrey D. "Artificial Intelligence 623 lecture notes", 2013.

[16] Clark, Jeffrey D. "Artificial Intelligence 823 lecture notes", 2013.

[17] Dash, Manoranjan and Huan Liu. "Feature selection for classification". *Intelligent data analysis*, 1(1-4):131–156, 1997.

[18] Dubois, Janie, Jean-Claude Wolff, John K Warrack, Joseph Schoppelrei, and E Neil Lewis. "NIR chemical imaging for counterfeit pharmaceutical products analysis". *SPECTROSCOPY-SPRINGFIELD*, 22(2):40, 2007.

[19] Duda, Richard O, Peter E Hart, and David G Stork. "Pattern Classification and Scene Analysis 2nd ed." 1995.

[20] Enzweiler, Markus and Dariu M Gavrila. "Monocular pedestrian detection: Survey and experiments". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2179–2195, 2009.

[21] Farnsworth, "The voice of 12", Clare.

[22] Frankel, Kevin, Sonia Sousa, Rick Cowan, and Melanie King. *Concealment of the warfighter's equipment through enhanced polymer technology*. Technical report, DTIC Document, 2004.

[23] Grimm, David C, David W Messinger, John P Kerekes, and John R Schott. "Hybridization of hyperspectral imaging target detection algorithm chains". *Defense and Security*, 753–763. International Society for Optics and Photonics, 2005.

[24] Guo, Yanlin, Steve Hsu, Ying Shan, Harpreet Sawhney, and Rakesh Kumar. "Vehicle fingerprinting for reacquisition & tracking in videos". *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, 761–768. IEEE, 2005.

[25] Guyon, Weston and Barnhill. "Gene Selection for Cancer Classification using Support Vector Machines". *Machine Learning*. 2002.

[26] Hammer, Barbara and Thomas Villmann. "Generalized relevance learning vector quantization". *Neural Networks*, 15(8):1059–1068, 2002.

[27] Han, Jun and Claudio Moraga. "The influence of the sigmoid function parameters on the speed of backpropagation learning". *From Natural to Artificial Neural Computation*, 195–201. Springer, 1995.

[28] Haran, Terence. "Short-wave infrared diffuse reflectance of textile materials". *Physics & Astronomy Theses*, 5, 2008.

[29] Haykin, Simon S, Simon S Haykin, Simon S Haykin, and Simon S Haykin. *Neural networks and learning machines*, volume 3. Prentice Hall New York, 2009.

[30] Hersey, melvin and Culpepper. "Dismount Modeling and Detection from Small Aperture Moving Radar Platforms". *IEEE*. 2008.

[31] Herweg, Jared, John Kerekes, and Michael Eismann. "Hyperspectral imaging phenomenology for the detection and tracking of pedestrians". *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, 5482–5485. IEEE, 2012.

[32] Herweg, Jared A, John P Kerekes, Emmett J Ientilucci, and Michael T Eismann. "Spectral variations in HSI signatures of thin fabrics for detecting and tracking of pedestrians". *SPIE Defense, Security, and Sensing*, 80400G–80400G. International Society for Optics and Photonics, 2011.

[33] Hrechak, Andrew K and James A McHugh. "Automated fingerprint recognition using structural matching". *Pattern Recognition*, 23(8):893–904, 1990.

[34] Incorporated, ASD. "FieldSpec 3 Hi-Res Portable Spectroradiometer". URL http://www.asdi.com/products/fieldspec-3-hi-res-portablespectroradiometer.

[35] Jin, Xiaoying, Scott Paswaters, and Harold Cline. "A comparative study of target detection algorithms for hyperspectral imagery". *SPIE Defense, Security, and Sensing*, 73341W–73341W. International Society for Optics and Photonics, 2009.

[36] John, George H, Ron Kohavi, Karl Pfleger, et al. "Irrelevant Features and the Subset Selection Problem." *ICML*, volume 94, 121–129. 1994.

[37] John, George H, Ron Kohavi, Karl Pfleger, et al. "Irrelevant features and the subset selection problem". *Proceedings of the eleventh international conference on machine learning*, volume 129, 121–129. San Francisco, 1994.

[38] Jolliffe, Ian. *Principal component analysis*. Wiley Online Library, 2005.

[39] Kay, Steven M. *Fundamentals of Statistical signal processing, Volume 2: Detection theory*. Prentice Hall PTR, 1998.

[40] Khazai, Safa, Saeid Homayouni, Abdolreza Safari, and Barat Mojaradi. "Anomaly detection in hyperspectral images based on an adaptive support vector method". *Geoscience and Remote Sensing Letters, IEEE*, 8(4):646–650, 2011.

[41] Kim, Z and Jitendra Malik. "Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking". *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, 524–531. IEEE, 2003.

[42] Kira, Kenji and Larry A Rendell. "The feature selection problem: Traditional methods and a new algorithm". *Proceedings of the National Conference on Artificial Intelligence*, 129–129. John Wiley & Sons Ltd, 1992.

[43] Kohavi, Ron and George H John. "Wrappers for feature subset selection". *Artificial intelligence*, 97(1):273–324, 1997.

[44] Kohonen, Teuvo. *Self-organizing maps*, volume 30. Springer Verlag, 2001.

[45] Koller, Daphne and Mehran Sahami. "Toward optimal feature selection". 1996.

[46] Kubik, Maria. "Hyperspectral imaging: a new technique for the non-invasive study of artworks". *Physical Techniques in the Study of Art, Archaeology and Cultural Heritage*, 2:199–259, 2007.

[47] Kumar, Ajay. "Computer-vision-based fabric defect detection: a survey". *Industrial Electronics, IEEE Transactions on*, 55(1):348–363, 2008.

[48] Kumar, Ajay and Grantham KH Pang. "Defect detection in textured materials using Gabor filters". *Industry Applications, IEEE Transactions on*, 38(2):425–440, 2002.

[49] Langley, Pat et al. *Selection of relevant features in machine learning*. Defense Technical Information Center, 1994.

[50] Liu, Ce, Lavanya Sharan, Edward H Adelson, and Ruth Rosenholtz. "Exploring features in a bayesian framework for material recognition". *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 239–246. IEEE, 2010.

[51] Liu, Huan and Lei Yu. "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution". *Correlation-Based Filter Solution". In Proceedings of The Twentieth International Conference on Machine Leaning (ICML-03)*, 856–863. ICM, Washington, D.C., 2003.

[52] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. *The WEKA Data Mining Software*. The University of Waikato, 2009.

[53] Martin, Richard. "Estimation and Detection EENG663 lecture notes", 2013.

[54] Mason, Mark T and Isaiah Coleman. *Study of the Surface Emissivity of Textile Fabrics and Materials in the 1 to 15MU Range*. Technical report, DTIC Document, 1967.

[55] Nunez, Abel S and Michael J Mendenhall. "Detection of human skin in near infrared hyperspectral imagery". *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, volume 2, II–621. IEEE, 2008.

[56] Okamoto, Hiroshi and Won Suk Lee. "Green citrus detection using hyperspectral imaging". *Computers and Electronics in Agriculture*, 66(2):201–208, 2009.

[57] Rangaswamy, Muralidhar, JS Goldstein, Michael L Piccolo, and Jacob D Greisbach. "Detection of Dismounts using Synthetic Aperture Radar", 2010.

[58] Roller, D, Kostas Daniilidis, and Hans-Hellmut Nagel. "Model-based object tracking in monocular image sequences of road traffic scenes". *International Journal of Computer Vision*, 10(3):257–281, 1993.

[59] Ruiz, LA, A Fdez-Sarría, and JA Recio. "Texture feature extraction for classification of remote sensing data using wavelet decomposition: a comparative study". *20th ISPRS Congress*. 2004.

[60] Shaw, Gary and Dimitris Manolakis. "Signal processing for hyperspectral image exploitation". *Signal Processing Magazine, IEEE*, 19(1):12–16, 2002.

[61] Soliman, Neil A. *Hyperspectral-Augmented Target Tracking*. Technical report, DTIC Document, 2008.

[62] Stein, David WJ, Scott G Beaven, Lawrence E Hoff, Edwin M Winter, Alan P Schaum, and Alan D Stocker. "Anomaly detection from hyperspectral imagery". *Signal Processing Magazine, IEEE*, 19(1):58–69, 2002.

[63] Wildes, Richard P. "Iris recognition: an emerging biometric technology". *Proceedings of the IEEE*, 85(9):1348–1363, 1997.

[64] Zhang, Lefei, Liangpei Zhang, Dacheng Tao, and Xin Huang. "On combining multiple features for hyperspectral remote sensing image classification". *Geoscience and Remote Sensing, IEEE Transactions on*, 50(3):879–893, 2012.

[65] Zhang, Xudong, Nicolas H Younan, and Charles G O'Hara. "Wavelet domain statistical hyperspectral soil texture classification". *Geoscience and Remote Sensing, IEEE Transactions on*, 43(3):615–618, 2005.

[66] Zhou, Jian, Dimitri Semenovich, Arcot Sowmya, and Jun Wang. "Sparse Dictionary Reconstruction for Textile Defect Detection". *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 1, 21–26. IEEE, 2012.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
27–03–2014	Master's Thesis	Oct 2012–Mar 2014

4. TITLE AND SUBTITLE

Textile Fingerprinting for Dismount Analysis in the Visible, Near, and Shortwave Infrared Domain

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Yeom, Jennifer S., Second Lieutenant, USAF

5d. PROJECT NUMBER

14G282

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-14-M-86

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Labratory
711th Human Performance Wing
Lead Scientist, Dr. Darrell F. Lochtefeld
2800 Q Street, Bldg 824
Wright Patterson AFB, OH 45433
darrell.lochtefeld@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/RHXBA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The ability to accurately and quickly locate an individual, or a dismount, is useful in a variety of situations and environments. A dismount's characteristics such as their gender, height, weight, build, and ethnicity could be used as discriminating factors. Hyperspectral imaging (HSI) is widely used in efforts to identify materials based on their spectral signatures. More specifically, HSI has been used for skin and clothing classification and detection. The ability to detect textiles (clothing) provides a discriminating factor that can aid in a more comprehensive detection of dismounts. This thesis demonstrates the application of several feature selection methods (*i.e.,* support vector machines with recursive feature reduction, fast correlation based filter) in highly dimensional data collected from a spectroradiometer. The classification of the data is accomplished with the selected features and artificial neural networks. A model for uniquely identifying (fingerprinting) textiles are designed, where color and composition ard determined in order to fingerprint a specific textile. An artificial neural network is created based on the knowledge of the textile's color and composition, providing a uniquely identifying fingerprinting of a textile. Results show 100% accuracy for color and composition classification, and 98% accuracy for the overall textile fingerprinting process.

15. SUBJECT TERMS

dismount detection, textile fingerprinting, feature selection, classification

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			LtCol Jeffrey D. Clark, PhD (ENG)
U	U	U	UU	96	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636x4614 Jeffrey.Clark@afit.edu